

Document Management, Document Imaging



IAMS LIC 2000 Conference Tides of Technology

Workshop: Document Management and Document Imaging for Aquatic and Marine Libraries

September 30 & October 1, 2000, Victoria, British Columbia, Canada

Presented by: SteveGilheany@ArchiveBuilders.com

www.ArchiveBuilders.com 24002-019

A-1

International Association of Aquatic and Marine Science Libraries and Information Centers

IAMS LIC 2000 Conference: Tides of Technology

Workshop: Document Management and Document Imaging for Aquatic and Marine Libraries

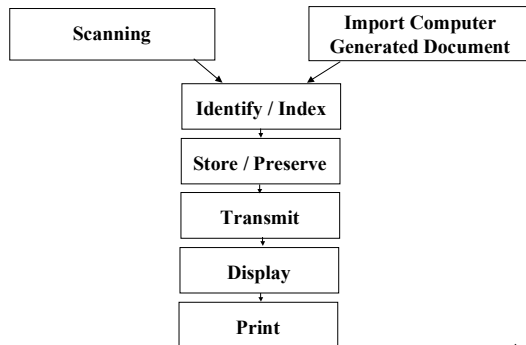
September 30 & October 1, 2000, Victoria, British Columbia, Canada

Presented by: Steve Gilheany Tel: +1(310) 937-7000 Fax: +1(310) 937-7001
1209 Manhattan Avenue, C-14, Manhattan Beach, California 90266
<http://www.ArchiveBuilders.com> SteveGilheany@ArchiveBuilders.com

www.ArchiveBuilders.com 24002-019

A-2

Flow



www.ArchiveBuilders.com 24002-019

A-3

Daily Schedule

Day 1

- 1 - A
- Break
- 1 - B
- Lunch
- 1 - C
- Break
- 1 - D

Day 2

- 2 - A
- Break
- 2 - B
- Lunch
- 2 - C
- Break
- 2 - D

www.ArchiveBuilders.com 24002-019

A-4

Day/Part Slide Topic

Day/Part	Slide	Topic
1 - A	A - 1	First Principles / Overview
1 - B	A - 59	Scanning / Printing
1 - C	B - 7	Importing / Preserving Files
1 - D	B - 186	Index / Search
2 - A	B - 203	Store / Purge / Preserve / Display
2 - B	B - 291	Internet / Systems
2 - C	C - 109	RFP / Records Management
2 - D	C - 209	Paradigms, References, & Future

www.ArchiveBuilders.com 24002-019

A-5

Overhead Transparencies

1. Raster Scanned Characters
2. One-to-One Pixel Image and Original Size Image
3. Mechanism of a Low Cost Laser Printer
4. Crafted Edge Character, 'O'
5. Crafted Edge Character, 'P'
6. Typographic Onomatopoeia
7. Spiral Infinity
8. Three Types of Soft Copy Display
9. Halftone Screen and Macropels
10. Screened Image
11. Rescreened Image
12. Rescaling, 14. Nyquist 2X Sampling
16. Network Diagram

www.ArchiveBuilders.com 24002-019

A-6

Document Management, Document Imaging

White Papers (Handouts) (1) at:

<http://www.ArchiveBuilders.com/abpapers.html>

1. **Microsoft Evolution: The 3.1 Flavors of Windows 2000 Become the Microsoft Environment**
2. Discussing DVDs Applications (1)
3. Sizing a Document Management System: Image Size Estimates: All Types of Digitized Documents (2)
4. The Decline of Magnetic Disk Storage Cost Over the Next 25 Years (3)
5. Document Management System Search Features: The More, The Merrier
6. Paper Sizes and Paper Weight: Metric and US Standards
7. COLD, COOL, COM, Greenbar, and Your Bank Statement
8. The Use of Future Digital Data Sources in Land Use Planning Documents
9. **Computer Storage Requirements for Various Digitized Document Types (2)**

www.ArchiveBuilders.com 24002-019

A-7

White Papers (Handouts) (2) at:

<http://www.ArchiveBuilders.com/abpapers.html>

10. Preserving Information Forever and a Call for Emulators
11. Projecting the Cost of Magnetic Disk Storage Over the Next 10 Years (3)
12. The DVD in Libraries (1)
13. Computer Technology in Libraries (1)
14. **The Next 3 Years on the Internet**
15. Microsoft Initiatives: Day by Day, 1996 to the Present
16. Evolution of Intel Microprocessors: 1971 to 2002 (4)
17. Moore's Law and Knowledge Management (4)
18. Fascicular Printing, Composing and Printing One Page at a Time, Like Laser Printing

www.ArchiveBuilders.com 24002-019

A-8

White Papers (Handouts) (3) at:

<http://www.ArchiveBuilders.com/abpapers.html>

19. Email, Voicemail, and Web Page Appropriateness Reminder for Employees
 20. Disaster Plan
 21. How Digitizing Works
 22. How RAID (Redundant Array of Inexpensive magnetic Disks) Works
 23. Expected Useable Lifetime of Different Electronic Document Formats
 24. How the Internet Works: Internet Protocol Stack & Internet Structure
- Misc. Annotated Illustrations

Please note:

- (1) The contents of articles 002, 012 and 013 are similar.
- (2) 009 is a condensed version of 003.
- (3) 011 is a condensed version of 004.
- (4) 016 is a condensed version of 017.

www.ArchiveBuilders.com 24002-019

A-9

PDF Formats and Preserving Information Forever

- This is a new paper (22025)
- It includes updates to:
 - Average Image Sizes (22004 and 22009)
 - Magnetic Storage Costs (22003 and 22011)
 - DVDs (22002, 22012, and 22013)
 - Microprocessor History (22016 and 22017)
- Please check the website below for updates

www.ArchiveBuilders.com 24002-019

A-10

Goals of the Class

- To provide a first principals understanding of document management technology.
- To provide sufficient background to understand document management in the context of a marine or aquatic library.

www.ArchiveBuilders.com 24002-019

A-11

First Principles

www.ArchiveBuilders.com 24002-019

A-12

Document Management, Document Imaging

Perspective, Context

- Where will you be in 5, 10, 25 years?
- Where will your organization, technology, be in 5, 10, 25 years?
- Familiar tasks in which you excel are a blessing. Start with what you know.
- ----With a little help from your friends.
- These things do not automatically fit together.

www.ArchiveBuilders.com 24002-019

A-13

How Much Should You Plan to Learn About Computers?

- Everyone has an idea of what a well educated high school graduate should know.
 - Not someone who just made it, **well educated**.
- Everyone agrees that such a high school graduate should have an understanding of computers.
 - **This** level of knowledge could be your goal.

www.ArchiveBuilders.com 24002-019

A-14

3 Years Experience

- Would you hire a manager without 3 years experience?
 - What do you need to know to get the interview?
 - What do you need to know to get hired?
- How will you get 3 years experience before you commit your organization to imaging?
 - Start now.

www.ArchiveBuilders.com 24002-019

A-15

Pitfalls

- Learning what is going away
- Learning too much about too little
 - (a detail, a nit, minutiae)
- Learning about something you will never use
- Learning in a way that teaches you to hate learning
- The above are considered in selecting materials for this presentation

www.ArchiveBuilders.com 24002-019

A-16

Learning: Insight, Stability

- Three points of view or three experiences provide stability and insight.
- After driving three cars, you can drive any car.
- Milk stools have three legs.

www.ArchiveBuilders.com 24002-019

A-17

Learning Sequence

- Points of Light
- Let a thousand flowers bloom
- Draw Lines, Fill In the Outlines
- Deepen, Res-Up, Enrich, add Color
- Synthesize
- Holographic / It just is

www.ArchiveBuilders.com 24002-019

A-18

Document Management, Document Imaging

Jargon

- We use words to communicate.
- Mutual agreement on the meaning of words is the first step.
- Any word can be used as Jargon.
- Any word can be used to communicate.
- Intent makes the difference.

www.ArchiveBuilders.com 24002-019

A-19

DWIM (1)

- DWIM (Do What I Mean)
- DWIM is the solution to:
 - “I would have told you if you had asked the right question.”
- Everyone is specialized.
 - Our knowledge and experience sets just barely overlap.
 - We are forming a human chain, stretched across the ice to save someone who has fallen into the water.
 - We are muddling through until digital document management becomes just a part of the world — until it just is.

www.ArchiveBuilders.com 24002-019

A-20

DWIM (2)

- Has this happened before?
 - Almost everything has happened before!
- Real steam locomotives are always on the edge of exploding.
 - It is only the skill of the engineer, who is really an engineer, that keeps the steam locomotive running safely.
- Document Management and Document Imaging are still in the engineering stage.
 - We live in interesting times

www.ArchiveBuilders.com 24002-019

A-21

New Technology Today (1)

- DVD (commonly *D*igital *V*ideo *D*isk)
 - 17 GBytes ROM, 7.4 WORM, 5.2 RAM
- 73 GigaByte 3 1/2 inch magnetic disk drive:
 - 500 boxes stored (62.5 GByte x 8 boxes per GigaByte = 400 boxes)
 - 50 boxes per minute (100 MegaBytes per second)
 - 10,000 RPM, 6.5 millisecond access time
 - 10 G operating shock, 2 pounds, 1/50 cu. ft. (.62 liter)

www.ArchiveBuilders.com 24002-019

A-22

New Technology Today (2)

- DWDM (*D*ense *W*avelength *D*ivision *M*ultiplexing)
 - Same fibers, new transmitters and receivers
 - 320 Gigabits per second (320 boxes per second) (80 file cabinets per second)
- Xerox: new 1,300 page per minute laser printer
 - 1/2 box per minute

www.ArchiveBuilders.com 24002-019

A-23

Vanilla System - Standalone Desktop

- On a PC Desktop
 - 50 GigaBytes Hard Drive
 - DVD Burner (WORM and RW) (*W*rite *O*nce *R*ead *M*any, *R*ead *W*rite)
- 1 Million Scanned Pages (100 file cabinets) (400 boxes)

www.ArchiveBuilders.com 24002-019

A-24

Document Management, Document Imaging

Vanilla System - Server

- On a PC Server
 - 120 x 73 GigaByte Hard Drives
 - 8 TeraBytes
 - on Fiber Channel in one 19 inch (wide) rack
 - RAID 3 (not striped), mirrored (RAID Level 1) to DVD
 - (Redundant Array of Inexpensive Disks)
 - (commonly Digital Video Disk)
 - 80 MegaBytes per second (1,000 pages per second)
 - DVD burner (to mirror to DVD)
 - 160 Million Scanned Pages (16 thousand file cabinets) (64 thousand boxes)

www.ArchiveBuilders.com 24002-019

A-25

Renaissance Person

- To know what happens between striking a key on the keyboard and a letter appearing on a PC screen requires a Masters Degree!
 - Today
- If you took the 6 years required to get a Masters Degree, you would not succeed in learning what happens, because in 6 years it will not be possible for a single person to know what happens!

www.ArchiveBuilders.com 24002-019

A-26

Where Does It Fit

- Wisdom
- Knowledge
- Information
- Data
- Bits
- Discernable differences

www.ArchiveBuilders.com 24002-019

A-27

Focus of the Presentation

- 1 Bit Pixels
- 50K Bytes per Page
- 8 1/2 by 11 Images
- 10K Pages per File Cabinet
- 200 & 300 dpi
- 2 File Cabinets = 1 GByte

www.ArchiveBuilders.com 24002-019

A-28

Scanned Objects

- This course does not deal with:
 - Physical Objects
 - Parts of Assemblies
 - Database Entries (Except as Metadata)

www.ArchiveBuilders.com 24002-019

A-29

Images

- Pages
- Documents
- Multimedia
- Hypertext / Internet

www.ArchiveBuilders.com 24002-019

A-30

Document Management, Document Imaging

Hype

- Document Imaging (Flashy)
- Multimedia (Flashier)
- Internet (Flashiest) (And a Good Idea)
- Therefore:
Document Imaging on the Internet

www.ArchiveBuilders.com 24002-019

A-31

There are no right answers.

www.ArchiveBuilders.com 24002-019

A-32

There Are No Right Answers

- Not only do people want right answers . . .
- They want as many right answers as possible.

www.ArchiveBuilders.com 24002-019

A-33

There Are No Right Answers

- A design based on a 'right answer' will fail.
- You must fully understand your needs and resources to design a system.
- At each point in time you must use judgement to choose a course of action.
 - Right answers are based on a past reality.
- Past decisions have weight as past decisions, but current actions must be based on current decisions that take into account all current circumstances.

www.ArchiveBuilders.com 24002-019

A-34

Do You Have to Be Technical?

www.ArchiveBuilders.com 24002-019

A-35

What is the Best Way to Design a System Backup?

- Imagine explaining to each person in your organization that you have lost all your records.
 - You will find the people you need to do it right.

www.ArchiveBuilders.com 24002-019

A-36

Document Management, Document Imaging

Start Anywhere

- This presentation is for non-technical managers.
 - Leadership is the one essential needed.
- There is very little dependency of one part of this presentation on another.
 - If you do not understand it, ignore it.
- You do not have to understand the technical details to benefit.
 - If you miss it, it will not matter.

www.ArchiveBuilders.com 24002-019

A-37

For Each Detail There is a Higher Level Managerial Perspective

- This limits technical dependencies.
 - If grayscale is not clear
 - Remember pixels
 - If pixels and compression are not clear
 - Remember 50 KiloBytes per page
 - If KiloBytes sound like MegaBytes and GigaEtc.
 - Remember 1 file cabinet per CD (Compact Disc)

www.ArchiveBuilders.com 24002-019

A-38

Mathematics

- This course is designed for high school graduates who avoided mathematics in high school.
- The course makes reference to counting and arithmetic. Your indulgence is requested.

www.ArchiveBuilders.com 24002-019

A-39

Commercial Products

- Commercial products are mentioned only in support of educational objectives.
- Commercial products are mentioned only to the extent required for the educational objective.
- Vendor support is gratefully acknowledged.

www.ArchiveBuilders.com 24002-019

A-40

Other Vendors

- Microsoft.com, Adobe.com, ESRI.com
 - There are many vendors that have complete product lines that compete with these vendors.
 - For example, it is not true that the US Justice department was compelled to use Microsoft Office to prosecute the Microsoft anti-trust case, as the US Justice department has publicly stated.
- One vendor is selected as an example
 - The one most likely for students to have or acquire
 - The need for users to choose quickly is one of the market leaders' greatest barriers to entry

www.ArchiveBuilders.com 24002-019

A-41

Document Management Overview

www.ArchiveBuilders.com 24002-019

A-42

Document Management, Document Imaging

Document Management

- Media (What is it on?)
- Metadata (Where did it come from?)
- Document (It) (Can it be digitized?)
 - Lord Kelvin: “It only matters if it can be measured.”
 - Corollary: “If it can be measured it has nothing to do with equality.”
- Index (Where is it?)
- Links (References to and from?)
 - Creating the ‘one worldwide document’

www.ArchiveBuilders.com 24002-019

A-43

Management Sequence

- Select / Create
- Capture
- Store
- Identify (Metadata)
- Index
- Link (Added by Users Over Time)
- View / Print / Send
- Preserve / Destroy

www.ArchiveBuilders.com 24002-019

A-44

The Contexts of Document Management

www.ArchiveBuilders.com 24002-019

A-45

Why So Many Contexts of Document Management?

- Why study so many types of documents?
 - Computer techniques are developed for all types of documents simultaneously.
 - Tools and techniques developed for one type of documents are often useful for other types of documents.
 - Document managers may be assigned to manage any type of documents.

www.ArchiveBuilders.com 24002-019

A-46

Documents (1)

- Accounting, Invoices
- Unstructured Text: Reports, Articles
- Active File Series: Medical Records, Customer Records
- Fax Input, Fax-Back Server
- Voice Mail, Email

www.ArchiveBuilders.com 24002-019

A-47

Documents (2)

- Remittance Processing: Credit Card Receipts
- Magazine and Book Production
- Oil Well Logs
- Cartography (maps)
- Cadastral (land) records

www.ArchiveBuilders.com 24002-019

A-48

Document Management, Document Imaging

Documents (3)

- MiddleWare (PC Document File Management)
 - Historical, based on a fix for 8.3 character file names
- Discovery Proceedings (Legal), Discoverability
- Engineering Document Management (CAD, Aperture Cards, roll size drawings)
- Public Record Documents in Government
- COLD, COOL, COM (Computer Reports)

www.ArchiveBuilders.com 24002-019

A-49

Documents (4)

- Records Management (Bankers Boxes)
- Microforms (All types of Microfilm)
- Archives, Library, Museum, Records Center
- Demand Printing (Replacing Offset Printing)
- Collateral Material (Advertising Literature)

www.ArchiveBuilders.com 24002-019

A-50

Documents (5)

- Structure Documents: Manuals (SGML)
- Document Objects
 - CAD (Computer Aided Design) Components
 - GIS (Geographic Information System) Entities
- Non-Linear Video and Audio Editing
 - Surveillance, “after the fact” person and object tracking in videos. Enhancement to 1984, the book
- Genealogy, Phylogenetics, Gene Sequencing

www.ArchiveBuilders.com 24002-019

A-51

Digital Library Examples

- U of Michigan, The Making of America
 - Un-deskewed Raster Images & OCR
 - at: <http://www.umdl.umich.EDU/moa/>
- UVA (Virginia)
 - Collects Electronic Documents (SGML)
 - at: <http://etext.lib.Virginia.EDU/>
 - Digital Library References (IFLA)
- International Federation of Library Associations
 - at: <http://www.nlc-bnc.ca/ifla/>
- Metadata (Dublin Core)
 - at: <http://www.dlib.org/dlib/february98/02weibel.html>

www.ArchiveBuilders.com 24002-019

A-52

Libraries and Other Document Management Stakeholders

www.ArchiveBuilders.com 24002-019

A-53

Libraries Favor Integrity Over Operational Access

- Access to information in libraries is not a critical component of an operational process
- Example: when an odor or liquid is leaking out of the ground in a city, the city’s substructure records must be accessed immediately
 - The system for storing substructure records is designed to meet this requirement
- Librarians must decide if the definition of information center includes records that carry this kind of liability and need for guaranteed access.

www.ArchiveBuilders.com 24002-019

A-54

Document Management, Document Imaging

Libraries Favor Access Over Preservation

- Documents have survived for long periods of time in libraries
- Libraries have special collections departments that function as archives
- Libraries transfer problem documents to archives so that funds can be allocated to acquisitions and reference services
- Libraries maintain a liaison with archives

www.ArchiveBuilders.com 24002-019

A-55

Libraries Favor Information Over Data, and Favor Knowledge Over Information

- Librarians attempt to save databases when all other options are unavailable
- Librarians do not have a database protection outreach program similar to the bibliographic instruction libraries provide
- Libraries prefer reports to collections of papers.

www.ArchiveBuilders.com 24002-019

A-56

Records Management Systems Exist to Preserve Information Forever (1)

- Non-procedural: no programs to preserve
 - Only record formats
 - (Avoid calculated cells in spreadsheets)
- Preserve a hierarchy of formats
 - Simplest: raster
 - provides a verification check on more complex formats
 - Vector (Outline fonts)
 - Native (Word processor)
 - More complex CAD and GIS formats
 - Computer Aided Design
 - Geographic Information Systems

www.ArchiveBuilders.com 24002-019

A-57

Records Management Systems Exist to Preserve Information Forever (2)

- Fascicles
 - Permanent, logical containers that facilitate preservation and backup
 - Digitally sealed electronic signatures
 - Paper 22025, Permanent Digital Records and PDF formats, describes a permanent records management system.

www.ArchiveBuilders.com 24002-019

A-58

Scan / Print

www.ArchiveBuilders.com 24002-019

A-59

Goals of the Session

- To provide a first principals understanding of digitizing.
 - What really happens to turn a document into ones and zeros and what really happens to turn ones and zeros into a document.
- To assist non-technical managers in participating in, and directing, technical discussions.
 - Managers should direct discussions.

www.ArchiveBuilders.com 24002-019

A-60

Document Management, Document Imaging

Doc Prep and Identification

- Many people start a project by buying a scanner.
 - That's all, just buying a scanner.
- Documents must be prepared to go through the scanner
 - And to go through the scanner at a reasonable speed so that the project will eventually be completed.
- It is important to know what you have and have not yet scanned in.
 - This question arises for many people after a large number of documents have been scanned and lost in the system.
 - Re-processing and rescanning the documents is often the only solution.

www.ArchiveBuilders.com 24002-019

A-61

Identification

- Not Indexing
- The minimum needed to keep the scanner going at full rated speed
- Used to track images in the system
 - Before the images have been gathered into document sub-objects
 - Before documents have been indexed.
- The minimum definition of Metadata

www.ArchiveBuilders.com 24002-019

A-62

Doc Prep

- Document Preparation
- Taking the staples out
 - & rubber bands, paper clips, postits, etc.
- Putting in document separator sheets
- Bates numbering (litigation support)
- Assembly into batches
 - For processing workflow
- De-Prep: preparing the documents for refileing

www.ArchiveBuilders.com 24002-019

A-63

Document Input Issues

- Simplex vs Duplex Scanners
- Mark Sense Separator Sheets
- Bar Code Data Input
- Bates Numbering
- 3, 10, 30, or 100 ppm
- 200, 300, 400, 1,000 dpi
- 8 1/2 by 11, Legal, E Size (48 by 36 inches)
 - Must handle older, slightly larger sizes

www.ArchiveBuilders.com 24002-019

A-64

Provenance for Doc Prep

- An archival term meaning the history of an object or document
- More prosaically (in working terms), in document imaging, it means:
 - Where were the staples before you prepped the documents for feeding through the scanner
 - Which folder was the document in, etc.
 - Which subsidiary's accounting department created the documents in the box labeled "accounting documents"

www.ArchiveBuilders.com 24002-019

A-65

Cost of Input

- 5 to 25 Cents per Page to Scan
- Digital Laser Copiers Include Free Input
- Need Digital Signatures and Seals for Word Processing Documents to Avoid Printing a File Copy
- Fax is in Electronic Form (If You Are Lucky, You Can Keep It that Way)

www.ArchiveBuilders.com 24002-019

A-66

Document Management, Document Imaging

What Do We Mean by

??????
~~~~ **Digitizing** ~~~~~  
??????

- If it is digital, it is good, it is better; well, it is *digital!*
- Do you know anyone who knows what digitizing is?
  - Can they explain it?
- Would digital toothpaste sell?

www.ArchiveBuilders.com 24002-019

A-67

## What Do We Mean by

??????  
~~~~ **Digitizing** ~~~~~  
??????

- What is a Digital Librarian (Archivist, Curator)?
 - Are you one? – Do you want to be one?
- What is a Digital Library (Archive, Museum)?
 - Do you have one? – Do you want one?

www.ArchiveBuilders.com 24002-019

A-68

What is a Raster

- The area where an image is produced on a kinescope (CRT – Cathode Ray Tube)
- Grid Definition
 - Size
 - Dimensionality: 1, 2, 3, or 4D
- Pixel Definition
- Compression
- Relatively Simple, Deterministic
- [Raster Image Slide]

www.ArchiveBuilders.com 24002-019

A-69

Raster Scanned Text -- ('AN')



www.ArchiveBuilders.com 24002-019

A-70

Raster Scanned Text -- ('FRACT')



www.ArchiveBuilders.com 24002-019

A-71

Glossary

- **Glyph** – the image of a character rendered in pixels.
- **Raster** – the scanned image created by a kinescope (a CRT, Cathode Ray Tube, such as that used in computer displays)
- **Pixel** – (PICTURE ELEMENTS) or pels (PICTURE ELEMENTS), an image sample area that is almost always square. Arranged in a grid, pixels form a raster image. A scanned page of a paper or microform document creates a digital image that is a raster of pixels.
- **Text image** – the content of a text record, often the contents of a page of text.

www.ArchiveBuilders.com 24002-019

A-72

Document Management, Document Imaging

Digital Imaging

www.ArchiveBuilders.com 24002-019

A-73

RGB and CMYK Units

- The RGB (Red, Green, Blue) color space uses units of grayscale that are based on A to D (Analog to Digital) (and D to A) conversion.
 - The RGB color space is based on the additive (generated) color used in computer display devices
- The CMYK (Cyan, Magenta, Yellow, Black) is similarly A to D (and D to A) based.
 - The CMYK color space is based on the subtractive (reflective) color used in printing
 - It is easy to convert between the RGB and the CMYK color spaces.
 - It is significantly more difficult to establish and convert to the color space created by phosphors on a display or ink on a medium with a moisture content in an environment with a changing relative humidity.

www.ArchiveBuilders.com 24002-019

A-74

Delta E (ΔE) Units

- The CIE L*a*b* (Commission Internationale de l'Eclairage (International Commission on Illumination) http://www.cie.co.at/cie/L*a*b* color space equation)
 - (Developed in 1976)
 - System uses the metrics of hue, saturation, and intensity
- Uses delta E (ΔE) units which are based on differences in color detectable by a person.
 - One delta E (ΔE) unit separates each pair of colors in the CIE L*a*b* system and by definition the separation of any two colors in the system is the minimum perceptible difference in two colors by a person at that point of separation in the color space.
 - It is possible to convert from the RGB or CMYK color spaces to the CLAB color space, but the metrics of the color spaces are relatively non-linear.

www.ArchiveBuilders.com 24002-019

A-75

More Than We Can See

- Because the RGB and CMYK color spaces are based on digitized color rather than perceived color, they are not optimized (compressed) to the minimum necessary number of colors required to reproduce a person's perception of color.
 - Therefore RGB and CMYK have more precision than is necessary, and can distinguish color differences that a person cannot distinguish (see)
 - Generally, the RGB and CMYK systems can reproduce all perceivable colors.
 - The RGB system can reproduce 16+ million colors and a person can see about 1 million colors.

www.ArchiveBuilders.com 24002-019

A-76

For Our Purposes

- The purpose of the following presentation is to provide some understanding of the method of converting:
 - From a color to a number in scanning
 - From a number to a color in printing
- The presentation is also intended to provide an understanding that an array of numbers, called pixels, represent a digital image.
 - The array of numbers, when reproduced visually is called a raster.
- For these purposes, the RGB, CMYK and CIE L*a*b* color spaces are equivalent.

www.ArchiveBuilders.com 24002-019

A-77

Our Visual Range

- We can only differentiate 256 (2^8) shades of gray.
- We can see about 1 million shades of gray at one time.
- Our visual range is about 1 billion shades of gray with a window of about 1 million shades available at any one time.
 - Our eyes adjust to sunlight
 - Our eyes adjust to a dark theater
 - Our visual ranges operate logarithmically, the way our hearing does, measured in decibels, and the way we perceive earthquakes, measured on the Richter Scale.

www.ArchiveBuilders.com 24002-019

A-78

Document Management, Document Imaging

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |

Chess Board as an Array of Numbers (Pixels)
0 = black
1 = white

www.ArchiveBuilders.com 24002-019

A-79

Pixels Do Not Have a Size

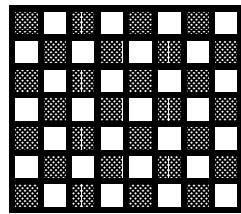
- Pixels are numeric values.
 - One numeric value for each component of the sample (e.g. Red, Green, Blue)
- Pixels have three attributes:
 - Shape (usually square)
 - The size of the area over which light was averaged to establish the pixel's numeric value.
 - The printed area over which a uniform color is reproduced that corresponds to the pixel's numeric value.

www.ArchiveBuilders.com 24002-019

A-80

Printed Chess Board

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |



Numbers (Pixels)

0 = Black 1 = White

www.ArchiveBuilders.com 24002-019

A-81

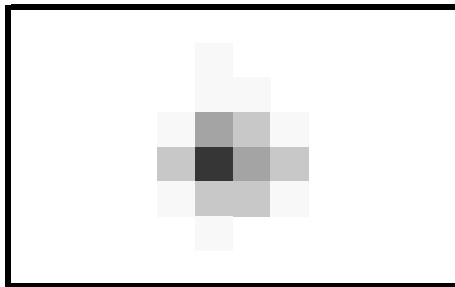
Quote from a JPL Researcher

- “We use 90 meter pixels.”
- Jet Propulsion Laboratory

www.ArchiveBuilders.com 24002-019

A-82

Headlines and Front Page Photo: New Planet (Star) (Galaxy) Discovered



Pixels of an Astronomical Size (simulated)

www.ArchiveBuilders.com 24002-019

A-83

Astronomy

- A single pixel can include:
 - An entire earth type planet (10 thousand kilometer pixels = 10 Mm, 10 Megameter pixel) (6 thousand miles)
 - A sun type star (1 million Kilometer pixels = 1 Gm, 1 Gigameter pixel), or a galaxy (100 thousand light year pixels = 1 Zm, 1 Zettameter pixel) (1/2 million miles)
 - The universe. The largest practical pixel is a 400 Ym, 400 Yottameter pixel, (250 billion trillion miles) the diameter of the observable universe.

www.ArchiveBuilders.com 24002-019

A-84

Document Management, Document Imaging

Chip Manufacturing

- Semiconductors are made using digital photographic techniques (pixels).
- Recently, microprocessor production processes were improved to .18 um (180 nm) (nanometer) design rules.
- The pixel size for semiconductor devices is now slightly less than 1/5 micron (200 nm). A micron is 1 / 1 millionth of a meter (40 millionths of an inch). (Semiconductor pixels are now about 8 millionths of an inch square.)
- This means that for the 200 nanometer (nm) pixels, and assuming 1/25th of the area was used for active transistors, a 1 millimeter (mm) square area (about the size of the head of a pin) could hold 25 MegaPels (25 million pixels) and 1 million transistors.

www.ArchiveBuilders.com 24002-019

A-85

Size Matters (Pixel Size that is)

- As transistors shrink:
 - Transistors take up less room so more fit on a chip
 - Computers can do more and remember more
 - Transistors have less inertia, so they switch faster
 - Computers go faster
 - Transistors have less inertia, so it takes less energy to switch them
 - Computer batteries last longer

www.ArchiveBuilders.com 24002-019

A-86

Summary of Pixels

- All pixels are identical in size and arrangement.
- All pixels are processed the same way.
- All pixels are scanned, displayed, and printed the same way.
- Each pixel has a location and a color.
 - Both given as numbers.
 - Location: latitude and longitude
 - Color: Amount of Red Green and Blue
 - Max on all 3 is white, minimum is black

www.ArchiveBuilders.com 24002-019

A-87

Pixels (Samples) Everywhere

www.ArchiveBuilders.com 24002-019

A-88

Bit-Mapped Images

- A bit-mapped image is a raster of pixels.
- Printed as a raster.
- Can be created by raster scanning.
- Can be created by a RIP (**R**aster **I**mage **P**rocessor) in a printer.

www.ArchiveBuilders.com 24002-019

A-89

Display Types

- All displays are now only raster
 - CRT (Cathode Ray Tube)
 - Laser Printer
 - Color Copier
 - Book
 - Fabric
 - Bread Wrapper (Packaging)
 - Truck Trailer (Billboards and Signage)

www.ArchiveBuilders.com 24002-019

A-90

Document Management, Document Imaging

The Grand Unification Theory

- Digitize Everything
- Use Pixels (Samples)
 - Make all Pixels the Same Size and Shape
 - Process all Pixels the Same Way

www.ArchiveBuilders.com 24002-019

A-91

Pixels

- Pixels, a Complete Definition
- Pixel Processing
- Pixels in a Page
- Pages in a System
- Systems Overview

www.ArchiveBuilders.com 24002-019

A-92

Samples

- Yes, everything can be represented as an array of samples.
- Pixels (2D)
- Pixels in Music and Other Times Series Data (1D)

www.ArchiveBuilders.com 24002-019

A-93

Pixels (Samples)

- 2D (Dimensional) (Square)
 - Fax, Photo, Documents, X-Rays
- 1D (Sound Volume)
 - Music, EKG's
- 3D (Voxels - Volume Elements)
 - MRI (Magnetic Resonance Imaging), Oil Pools, Aquifers
 - CNC (Computer Numerical Control)
 - Stereolithography
- 4D (3D + Time)
 - Weather, Pollution Plumes, Video Games
 - DNA Model Animations
- 2D + Time
 - Movies, MPEG Video Compression
 - (Motion Pictures Experts Group)

www.ArchiveBuilders.com 24002-019

A-94

Relationships

- Everything is more than related, --- "It just is!" (3 year olds have been telling us this for years.)
- Seeing relationships reduces the number of discrete things to learn
 - A rainbow is one octave

www.ArchiveBuilders.com 24002-019

A-95

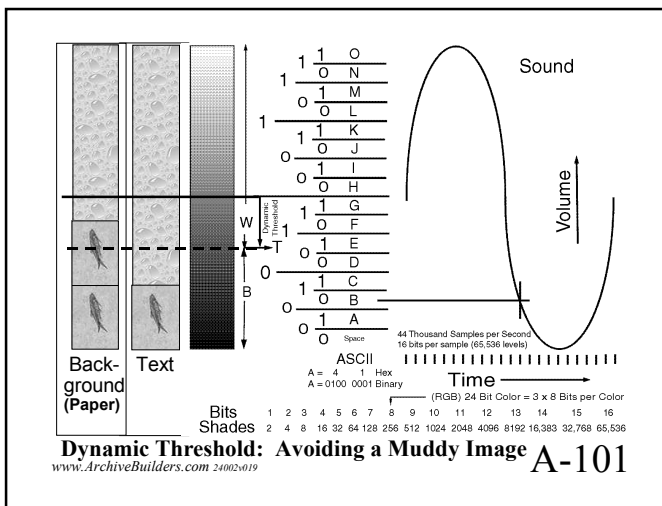
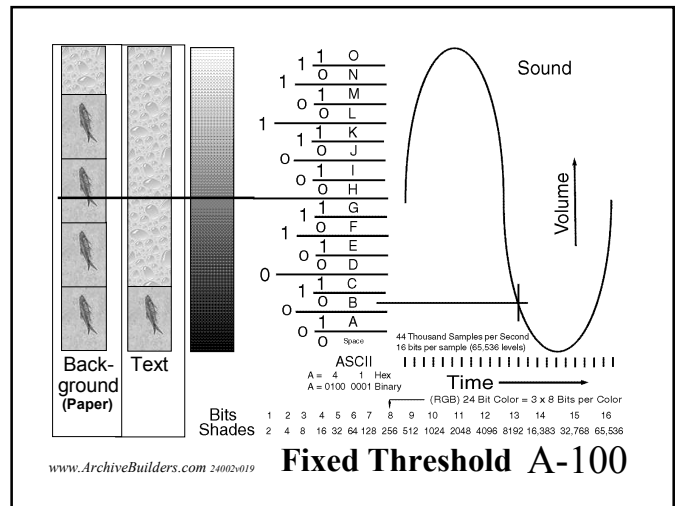
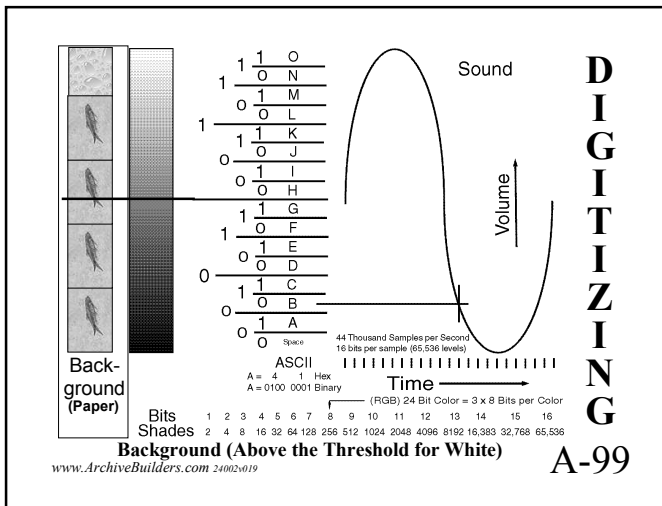
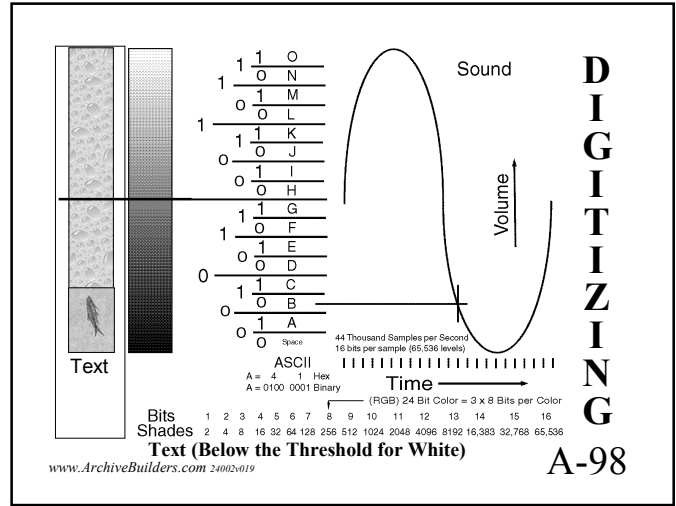
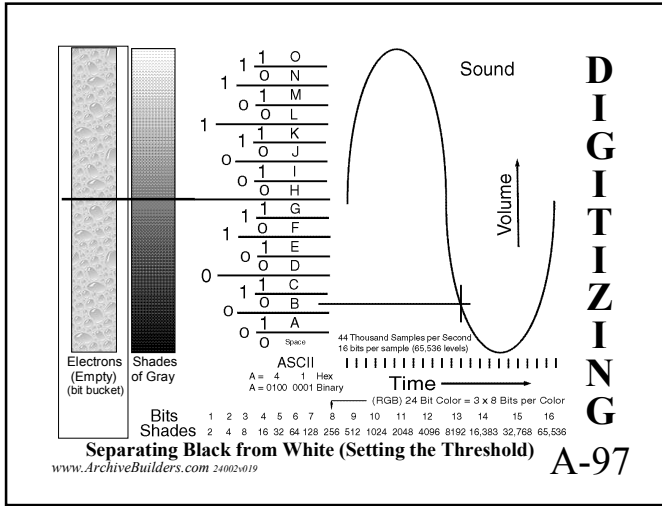
Pixels (Sampling)

- In Sequence [Presented on Blackboard]
 - Establish Sample Grid
 - Select Pixel to Sample
 - Sample the Pixel
 - Normalize the Sample
 - Quantify the Sample
 - Record

www.ArchiveBuilders.com 24002-019

A-96

Document Management, Document Imaging



Favorites in Computer Numerology

| | | | | | |
|-----|---------|--------|----------|----------------------------|----------|
| 2 | (2** 1) | 512 | (2** 9) | 131,072 | (2** 17) |
| 4 | (2** 2) | 1,024 | (2** 10) | 262,144 | (2** 18) |
| 8 | (2** 3) | 2,048 | (2** 11) | 524,288 | (2** 19) |
| 16 | (2** 4) | 4,096 | (2** 12) | 1,048,576 | (2** 20) |
| 32 | (2** 5) | 8,192 | (2** 13) | 2,097,152 | (2** 21) |
| 64 | (2** 6) | 16,384 | (2** 14) | 4,194,304 | (2** 22) |
| 128 | (2** 7) | 32,768 | (2** 15) | 8,388,608 | (2** 23) |
| 256 | (2** 8) | 65,536 | (2** 16) | 16,777,216 | (2** 24) |
| | | | | 4,294,967,296 | (2** 32) |
| | | | | 18,446,744,073,709,600,000 | (2** 64) |

NOTE: 264 IS APPROXIMATE IN EXCEL97, BUILD NUMBER NOT LISTED**

www.ArchiveBuilders.com 24002019

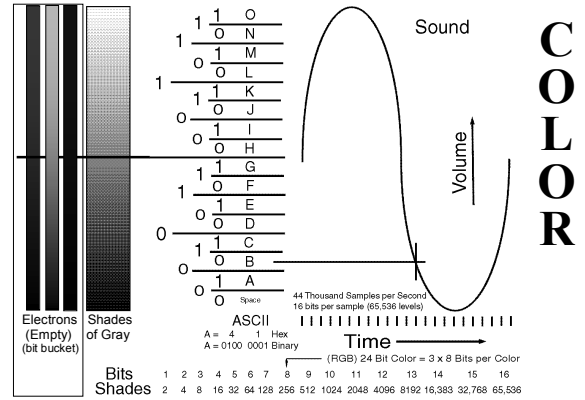
Document Management, Document Imaging

How Many Colors?

- With 8 bits of grayscale digitization, we can detect, represent, and reproduce $2^{*}8 = 256$ shades of red, green, or blue.
- For each of the 256 shades of red we can detect and reproduce, we can detect and reproduce 256 shades of green, giving us $256 * 256 = (2^{*}8) * (2^{*}8) = 65,536$ shades of red-green.
- For each of the 65,536 shades of red-green we can detect and reproduce, we can detect and reproduce 256 shades of blue, giving us $65,536 * 256 = (256 * 256) * 256 = (2^{*}8) * (2^{*}8) * (2^{*}8) = 2^{*}24 = 16,777,216$ shades of red-green-blue (full color)

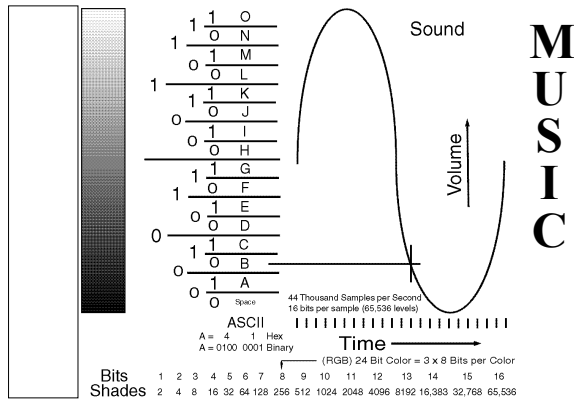
www.ArchiveBuilders.com 24002:019

A-103



24 Bit Color: 16,777,216 Colors ($2^{*}24$ Colors)
 = $(2^{*}8) * (2^{*}8) * (2^{*}8)$ Colors A-104

www.ArchiveBuilders.com 24002:019



16 Bit Music (Audio) Samples, 2 Samples per Unit Time for Stereo A-105

www.ArchiveBuilders.com 24002:019

Music: Connecting the Points of Light

- Music is sampled at about 50 KHz (50 thousand samples per sec.)
 - Actually: 44 KHz.
- Each sample is 16 bits (2 bytes, 65,536 levels) per channel
- Two channels (stereo)
- 2 bytes * 2 channels * 50 thousand = 200 thousand bytes per second
- 3,600 seconds per hour (60 seconds * 60 minutes)
- 3,600 seconds * 200 thousand bytes per second = 720 million bytes per hour
- Compares acceptably with 650 million bytes per 74 minute CD

www.ArchiveBuilders.com 24002:019

A-106

Feelings and Numbers

- Is it possible that the US educational system teaches our humanists to have a feeling of angst when they see blocks of numbers?
- Is it possible that the US educational systems teaches people who do well with numbers that feelings do not affect numerical solutions?
- We tear down expensive double-decker freeways in San Francisco.
 - Weren't they the most cost effective design?
 - But, aren't suspension bridges the most cost effective design?

www.ArchiveBuilders.com 24002:019

A-107

The Difference Between 12 Bit and 16 Bit Grayscale

- Nascent cancer tumor
 - With 12 bit grayscale X-Ray scanning, there are 4,096 detectable gradations of tissue density. The nascent cancer tumor tissue and the surrounding normal tissue are very close in density and fall in the same gradation, making the tumor undetectable. The doctor asks that you return in 6 months for a re-screening when the tumor is more detectable.
 - With 16 bit grayscale X-Ray scanning, there are 65,536 detectable gradations of tissue density. The nascent cancer tumor tissue and the surrounding normal tissue fall in adjacent gradations. The system stretches the image gamma, making the cancer tumor appear black and the surrounding tissue appear white. The doctor sees the tumor easily and removes the tumor before it can grow larger.

www.ArchiveBuilders.com 24002:019

A-108

Document Management, Document Imaging

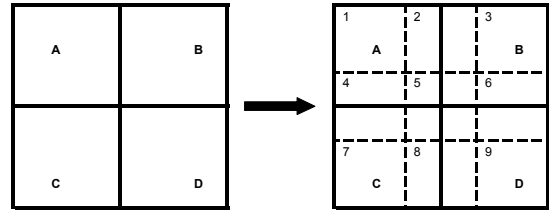
Multi-Spectral Imaging

- In color scanning, three color ranges are scanned: Red, Green, and Blue
 - These are segments of the electromagnetic spectrum (otherwise known as colors)
- The electromagnetic spectrum is very large, and segments can be digitized separately, just as we digitized Red, Green, and Blue separately.
 - The segments can be viewed separately, or combined in any number of different groups
 - We combine Red, Green, and Blue to see full color.
 - We assign the original RGB colors to the digitized RGB values, but RGB pseudo-color can be used to interpret any multi-spectral image

www.ArchiveBuilders.com 24002/019

A-109

Resize: Image Larger, Pixels Smaller, More Pixels, More Pseudo-Data

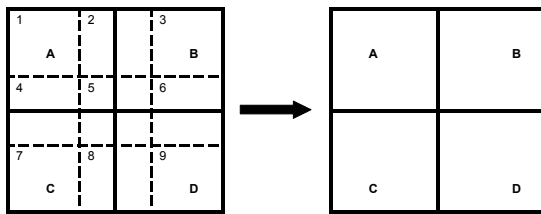


Creating 9 Pixels (1-9) where there were 4 Pixels (A-D)

www.ArchiveBuilders.com 24002/019

A-110

Resize: Image Smaller, Pixels Larger, Fewer Pixels, Less Pseudo-Data, Less Data



And, You Can Not Get Your Data Back Later

Creating 4 Pixels (A-D) from 9 Pixels (1-9)

www.ArchiveBuilders.com 24002/019

A-111

Moral:

- Archive your original images at their original scan resolution.
- Generate all derivative images from the archived images.
- Avoid having two versions of the archived image.
 - The two versions will always be different.
 - You will not be sure why they are different.
 - This is also true for data in databases

www.ArchiveBuilders.com 24002/019

A-112

Interpolation

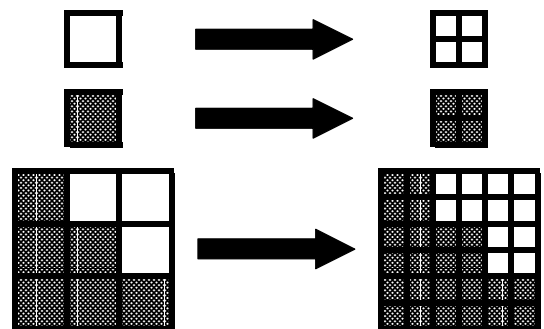
Making Up Data

(Better than extrapolation, but not much better.)

www.ArchiveBuilders.com 24002/019

A-113

Interpolation: Creating Pseudo-Data

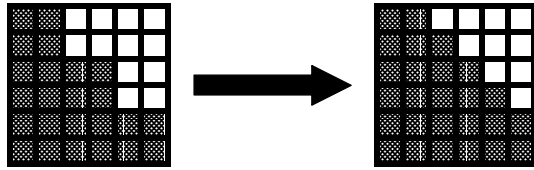


www.ArchiveBuilders.com 24002/019

A-114

Document Management, Document Imaging

Interpolation with Anti-Aliasing: More and Better Looking Pseudo-Data



www.ArchiveBuilders.com 24002-019

A-115

Scaling Out

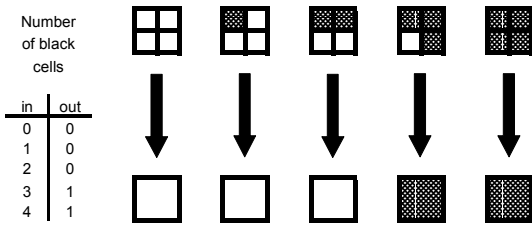
Throwing Away Data

(Better than interpolation, but not much better.)

www.ArchiveBuilders.com 24002-019

A-116

Scaling Out (Down), Eroding Line Art

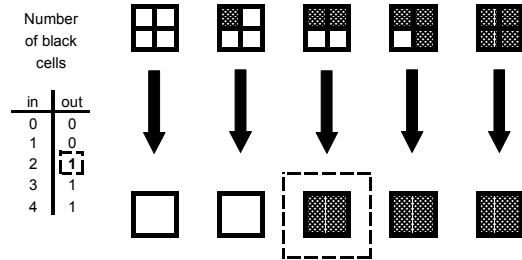


We Must Hang Together, Or We Shall Surely Hang Separately
From Many, One: E Pluribus Unum

www.ArchiveBuilders.com 24002-019

A-117

Scaling Out (Processing Down), Favoring Line Art



(But, this technique erodes line art in a negative image.)

www.ArchiveBuilders.com 24002-019

A-118

And Then There are Patterns to Attempt to Interpret or to Ignore

| Decimal | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------------|------|------|------|------|------|------|------|------|
| Hexadecimal | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Binary | 0000 | 0001 | 0010 | 0011 | 0100 | 0101 | 0110 | 0111 |

| | | | | | | | | |
|-------------|------|------|------|------|------|------|------|------|
| Binary | 1000 | 1001 | 1010 | 1011 | 1100 | 1101 | 1110 | 1111 |
| Hexadecimal | 8 | 9 | A | B | C | D | E | F |
| Decimal | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

www.ArchiveBuilders.com 24002-019

A-119

Anamorphic Resizing

- Shrinking (or enlarging) an image in one dimension more than another.
- Newspapers only shrink horizontally because they sell vertical inches of column space.
 - Newspaper do not want to shrink their papers vertically because it reduces their stock in trade, the column inch.

www.ArchiveBuilders.com 24002-019

A-120

Document Management, Document Imaging

Save Your Archival Images Quickly

- As images move around a system or organization, they may be unknowingly (surreptitiously) shrunk and enlarged, throwing away data.
- The longer images move around before being made the master read only image, the more likely that they will be compromised.
- Even if you are sure that the software is safe, the next release may not be safe, or a change in the setting may change the software's safe operation.

www.ArchiveBuilders.com 24002-019

A-121

OCR

www.ArchiveBuilders.com 24002-019

A-122

OCR

- OCR (**O**ptical **C**haracter **R**ecognition)
 - No longer based on optical processing
 - OCR algorithms process raster bit maps of pages
- ICR (**I**ntelligent **C**haracter **R**ecognition)
 - Has become synonymous with OCR
- 3D OCR
 - Uses greyscale information to improve character recognition of low resolution images (200 dpi)

www.ArchiveBuilders.com 24002-019

A-123

OCR Has Nothing to Do with any Particular Scanner

- OCR is done on a raster file.
- The raster file can be produced by any scanner.
- Settings on the scanner can affect the success of the OCR operation.
- The parameters for scanning and OCR may be in the same window, but this is just a convenience.
- The scanning and OCR operations are separate and occur in sequence.

www.ArchiveBuilders.com 24002-019

A-124

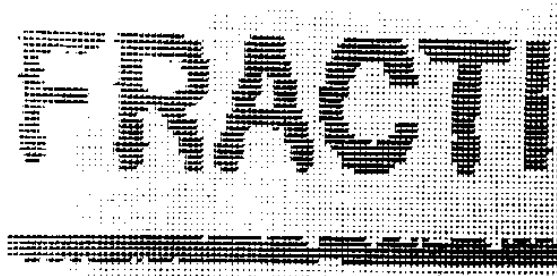
How OCR Works

- [At the Blackboard]
- How can characters that appear in a raster image be distinguished (one from another) by a computer program (algorithm).

www.ArchiveBuilders.com 24002-019

A-125

Raster Scanned Text -- ('FRACT')



www.ArchiveBuilders.com 24002-019

A-126

Document Management, Document Imaging

Uses of OCR

- Republishing documents
- Recovering lost copies of computer created documents
- Full text searching
- Barcode for data entry (built-in checksum)
 - Computers handle barcode as a font for printing
- Handwriting recognition requires pen tip pressure levels and order of pen strokes

www.ArchiveBuilders.com 24002-019

A-127

Required OCR Accuracy

- For full text recognition
 - Above 50 percent
 - Probably above 75 percent
- For republishing documents
 - Above 99.9 percent
 - This is 5 errors per 5,000 character page
 - (50 lines of 100 characters)

www.ArchiveBuilders.com 24002-019

A-128

Spotting Hits in the Document

- Identify the document containing the word
- Identify the page containing the word
- Identify the line containing the word
- Highlight the word in OCR text
- Highlight the word in the raster image
 - Adobe Acrobat

www.ArchiveBuilders.com 24002-019

A-129

Nyquist and Digital Artifacts

www.ArchiveBuilders.com 24002-019

A-130

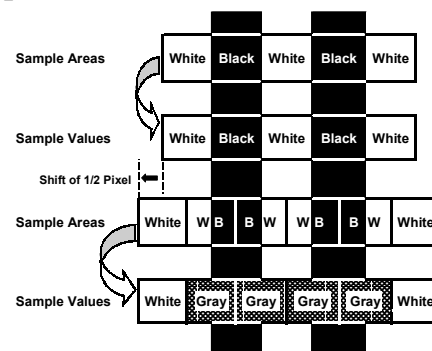
Nyquist 2X Over Sampling

- To resolve a given feature, such as a line, pixels one half the width of the feature or line must be used.
- This is because when a pixel covers an area that is half white and half black, the result is grey. Moving across an image of line pairs (using pixels that are the same width as the lines), the transition from black to white is recorded as the same grey as the next transition (which is from white to black) because they are both half black and half white.
- This results in the entire image being recorded as a uniform shade of grey.

www.ArchiveBuilders.com 24002-019

A-131

Sample Pixels = Line Width \Rightarrow All Gray



www.ArchiveBuilders.com 24002-019

A-132

Document Management, Document Imaging

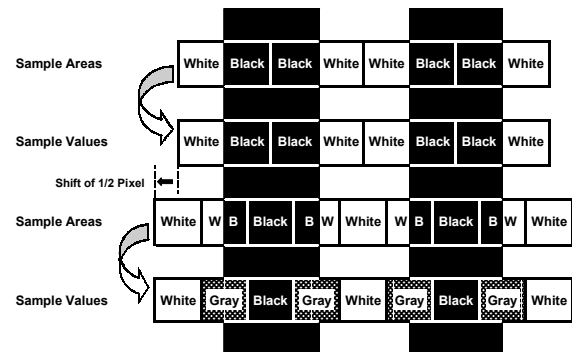
Pixel Registration

- Do your pixels line up with the lines of the line art on your images (document pages)?
 - Answer: sometimes yes, sometimes no
 - Some ‘yes’s and some ‘no’s on every page
- For 1/4 pixel accuracy in 300 dpi scanning, the paper must be positioned accurately to 1/1,200th inch
 - An impossibility
- Pixel registration is a random event
 - Not to be depended on

www.ArchiveBuilders.com 24002-019

A-133

Sample Pixels = 1/2 Line Width \Rightarrow Resolved



www.ArchiveBuilders.com 24002-019

A-134

Print vs. Scan Resolution

- A Given Print Resolution Quality is Equivalent to Twice the Nominal Linear Scan Resolution Quality.
 - By Nyquist's Sampling Theorem
- Examples: $\text{dpi (dots per inch)}$
 - 300 dpi scanned is equivalent to 150 dpi printed
 - 600 dpi scanned is equivalent to 300 dpi printed
 - 1,200 dpi scanned is equivalent to 600 dpi printed
 - 2,400 dpi scanned is equivalent to 1,200 dpi printed

www.ArchiveBuilders.com 24002-019

A-135

Nyquist and Generational Loss

- Generational loss when the image is redigitized
 - 1/2 of the resolution is lost in each dimension, leaving 1/2 of the resolution
 - The remaining areal resolution is $1/2 * 1/2$, or 1/4 of the previous generation's areal resolution
- Sampling Theorem contained in:
 - Harry Nyquist, "Certain Topics in Telegraph Transmission Theory," *Trans.*, AIEEE, Vol. 47, April 1928, pp. 617-644.

www.ArchiveBuilders.com 24002-019

A-136

Norsam and Generational Loss

- For some media, for example the Norsam Rosetta media (www.Norsam.com) (see Bibliography), the fascicular metadata can be written in such a way that it forms a readable miniature raster image on surface of the media that can be read with a microscope, eliminating the need for any computer software to decode the fascicle.
 - If the pixel registration of the miniature image can be recovered (areally synchronously) then there will be no Nyquist Sampling Theorem based generational loss of image resolution.

www.ArchiveBuilders.com 24002-019

A-137

Digital Artifacts: Halftoning and Moiré

- If the document being scanned was created in a way that created (or used) a grid of pixels, for example, a halftoning process, there can be a uniform misregistration of scanned pixels to printed pixels.
- This Misregistration results in an areal beat frequency between the areal frequency of the two pixel grids.
 - Just as two adjacent notes on a piano have a very low pitched beat tone.
- This areal beat frequency is seen as a moiré effect on the scanned image.
 - Related to the Nyquist sampling theorem
 - See PostScript Language Reference, Third Edition, in the bibliography, for details on half tone printing.

www.ArchiveBuilders.com 24002-019

A-138

Document Management, Document Imaging

Halftoning and Moiré

- A halftoned image
- A halftoned images with a Moiré interference pattern
- [At the blackboard]

www.ArchiveBuilders.com 24002-019

A-139

Barcode

www.ArchiveBuilders.com 24002-019

A-140

Paper File Management Advantages of Barcode

- Can not misread because it has an error correcting code.
- Easy to use, simplifies training
 - Can be wanded in either direction
 - Beeps when it reads correctly
- Reduces steps by checking file-outs on a computer.
- Misfile eliminated by wandling shelf barcode
- Eliminates lost-in-transit by wandling during in-transit transfers
- For computer printers, barcode is just a special font, like the 'Lil vehicles' font later in this presentation.

www.ArchiveBuilders.com 24002-019

A-141

Barcode

- Barcode can ease the transition from paper files to document imaging.
- Barcoded file folders and documents support scan on demand applications.
 - They integrate computer storage of scanned documents with computer management of stored boxes of paper.

www.ArchiveBuilders.com 24002-019

A-142

Barcoded Separator Sheets

- Barcoded document separator sheets are placed between documents that have been prepared for scanning (doc prepped).
 - The barcoded separator sheets are printed using data from existing document metadata databases.
 - These barcoded separator sheets include a barcoded key that is the link to the document metadata in an existing database.

www.ArchiveBuilders.com 24002-019

A-143

Mark Sense

- Color in a box or circle
- Used in multiple choice, compute scored, testing

www.ArchiveBuilders.com 24002-019

A-144

Document Management, Document Imaging

Compression

www.ArchiveBuilders.com 24002-019

A-145

Compression

- All images are stored and transmitted in compressed form.
- Compression can be non-destructive (Group 4) or destructive (some settings for JPEG).
 - JPEG (Joint Photographic Experts Group)
 - Destructive: changes the image to make it more compressible (and frequently cleaner and better looking).
 - Non-destructive: decompressed image is identical to original image (bit for bit) that was compressed.

www.ArchiveBuilders.com 24002-019

A-146

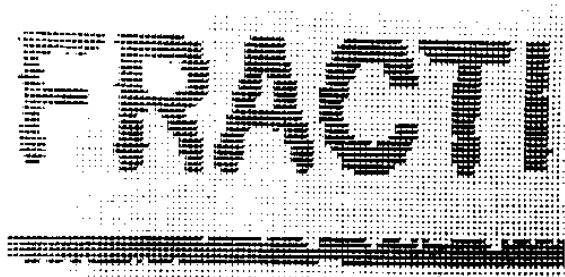
Compression

- [On Blackboard]
- With a 1 bit, black and white, raster image
 - G3 (Group 3 fax compression)
 - 1 dimensional (horizontal)
 - G4 (Group 4 fax compression)
 - 2 dimensional (horizontal and vertical)
 - The most common form of document image compression

www.ArchiveBuilders.com 24002-019

A-147

Raster Scanned Text -- ('FRACT')



www.ArchiveBuilders.com 24002-019

A-148

Compression / Compressing

- A 300 dpi image has about 10 million pixels
 - $(300 \times 300) \times (10 \times 10) = 9$ million
- A page image compresses to about 50 thousand bytes
 - 50 thousand bytes = 400 thousand bits (industry standard)
- Compression ratio is about 20 to 1
 - $8 \text{ million} / 400 \text{ thousand} = 20/1 = 20$
- A PC executes about 1 instruction per clock cycle (Hz)
 - A 1 GHz PC executes 1 billion instruction per second
- 100 instructions are available per pixel compressed / sec
 - $1 \text{ billion instructions} / 10 \text{ million pixels} = 100 \text{ inst./pixel}$

www.ArchiveBuilders.com 24002-019

A-149

Lossy / Lossless Compression

- Compression is lossless (non-destructive) if decompression reproduces the compressed file exactly, to the bit.
- JPEG (Joint Photographic Experts Group) compression has a lossless setting
 - It is almost impossible to insure that the lossless setting for any system is kept set to lossless
 - Many technical personnel see lossy compression as a big efficiency improvement, and are always trying to help

www.ArchiveBuilders.com 24002-019

A-150

Document Management, Document Imaging

Lossy Compression

- The process of scanning processes images is done in an inherently 'lossy way', (losslessness is a myth.)
- No digital image is ever the same as the original, by definition
 - Originals will continue to be scanned every 20 to 50 years until the originals disintegrate due to age.
- Grayscale thresholding, dynamic thresholding, and color gambit compression all affect image fidelity.
- It is the records manager's, archivist's, librarian's, or curator's professional seal on the digital file that says that the file is an accurate reproduction of the original.

www.ArchiveBuilders.com 24002-019

A-151

Specifying Degree of Compression Using Lossy Compression

- Do not specify the same compression ratio for all pages.
 - The noisy, dirty pages will be turned to mush because the compression will throw out the good information along with the noise. (Dirt on the image.) (Noise can be considered digital dirt.)
- Specify a degree of quality of image.
 - Compressed page files will vary in size, but the quality of the worst pages will be preserved as much as possible.
 - If the compressing system does not allow specifying a degree of quality, then set the ratio after viewing the worst pages, or, do not use the system.

www.ArchiveBuilders.com 24002-019

A-152

Quality Control

- 100 percent is best.
- All QC quality drifts over time.
- QC usually uses dedicated workstations.

www.ArchiveBuilders.com 24002-019

A-153

Scanners and Printers

- 4 little ones are faster than one big one
- 4 little ones are less expensive than one big one
 - Unless you consider labor
- 5-10 ppm (pages per minute)
- 20-30 ppm / 30-50 ppm / 100-200 ppm
- Simplex vs. Duplex

www.ArchiveBuilders.com 24002-019

A-154

Remittance Scanning

- Checks
 - Scanned front and back
 - Digital images available from banks on CD
- Credit card receipts
 - Soon only expenditures with transaction numbers (the exceptionally long numbers on credit card bills) will be allowed as tax deductions
- Up to 400 ips (*inches per second*)
 - 60 items per second

www.ArchiveBuilders.com 24002-019

A-155

Workflow

- Document Routing
- Structured vs Ad Hoc in Same System
- Proprietary vs Commercial Email
- Is Routing Setup Graphical?
- Can You Find In-Process Documents?
 - Even when you delete the node they are enqueued for?

www.ArchiveBuilders.com 24002-019

A-156

Document Management, Document Imaging

Archival vs. Web Quality

- Scan at highest quality
 - Archive images
- Process down to Web quality
 - Web quality images have very few other uses
 - Computer screen resolution
 - Windows that are smaller than full screen

www.ArchiveBuilders.com 24002-019

A-157

Summary of Pixels

- All pixels are identical in size and arrangement.
- All pixels are processed the same way.
- All pixels are scanned, displayed, and printed the same way.
- Each pixel has a location and a color.
 - Both given as numbers.
 - Location: latitude and longitude
 - Color: Amount of Red Green and Blue
 - Max on all 3 is white, minimum is black

www.ArchiveBuilders.com 24002-019

A-158

Moral

- Learn as much as you can about the processes your images and documents go through.
- Then, look at the finished product to make certain the effect of the processes is acceptable to you and your organization.

www.ArchiveBuilders.com 24002-019

A-159

Local Summary

- Everything can be digitized.
 - Most things start out in digital form
- The cost of digital storage is becoming a non-issue.
- Scanning is much more expensive than digital storage
 - The scanning process must be integrated with the normal workflow to make digital storage feasible

www.ArchiveBuilders.com 24002-019

A-160

How do you scan everything?

- “Scan everything.” is not the way to say “Put everything in the computer.”.
- Avoid planning to print material out for the purpose of scanning the material in.
- Almost everything starts in the computer today.
 - e.g. the best copy of a book is a typesetter file.
 - The problem is format conversion.

www.ArchiveBuilders.com 24002-019

A-161

End Runs: Eliminate the Need for Scanning

- You can not just scan everything in
 - There is just too much paper
- Everyone can not have a telephone
 - There would not be enough operators
 - Circa 1920

www.ArchiveBuilders.com 24002-019

A-162

Document Management, Document Imaging

End Runs: Eliminate the Need for Scanning

- Electronic signatures allow retention of electronic copy instead of printing a file copy
- Digital copiers automatically retain a copy
- Convert to the use of electronic forms
- Receive documents electronically
- Electronic transaction records
- COLD (Computer Output to Laser Disc)

www.ArchiveBuilders.com 24002-019

A-163

3D Scanning

- [Video of 3D scanning process]
- Scanning a scene to match digital movie matting to the physical background filmed.
- Object scanning
 - A cloud of points vs. a stereo photo pair

www.ArchiveBuilders.com 24002-019

A-164

Print

www.ArchiveBuilders.com 24002-019

A-165

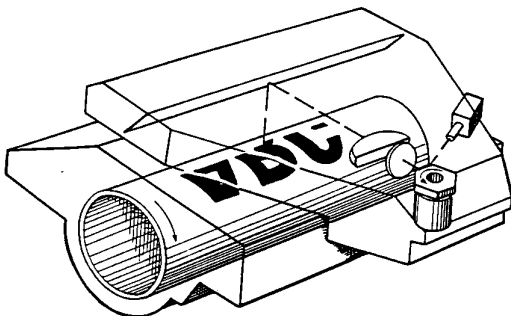
Printers Print Raster Images

- Sources of raster images:
 - Scanned raster bitmap
 - Printer generated bitmap from outline font
- [View the raster image produced by the computer projector.]

www.ArchiveBuilders.com 24002-019

A-166

Mechanism of a Laser Printer



www.ArchiveBuilders.com 24002-019

A-167

Mechanism of a Laser Printer (1)

- The dotted line in this picture represents a ray of light (laser beam) traveling from the laser to the drum of a laser printer.
- The light leaves the laser and is reflected from a rotating hexagonal (six sided) prism (mirror). The rotation is shown by the curved arrow.
- The laser beam is then reflected by a second, fixed, mirror, onto the drum of the laser printer.
- As the prism mirror rotates the laser beam moves across the drum from end to end. (This effect can be simulated and tested with a laser pointer and a small mirror.)

www.ArchiveBuilders.com 24002-019

A-168

Document Management, Document Imaging

Mechanism of a Laser Printer (2)

- The laser is turned on to reproduce black pixels and off to reproduce white pixels.
- The drum rotates (shown by a curved arrow) to place the laser printed rows of pixels side by side, painting a raster image on the drum.
- The drum is light sensitive and attracts (black writer) or does not attract toner (white writer) to areas exposed by the laser beam (when the beam is turned on).
- Toner is transferred to paper
- Toner is then fused to the paper with pressure and heat

www.ArchiveBuilders.com 24002-019

A-169

Print Quality

- 240 dpi: Traditional Mainframe Laser Printing
- 300 dpi: First Generation PC Laser Printers
- 600 dpi: Second Generation PC Laser Printers
- 1,200 dpi: Newspaper Halftone Printing
- 2,400 dpi: Medium Quality Halftone Printing
 - 2,540 dpi = 100 dpmm (dots per millimeter)
- 3,600 - 5,400 dpi High Quality Halftone Printing

www.ArchiveBuilders.com 24002-019

A-170

Raster Rendering Speed

- There is almost always an incompatibility between the raster image in the document management system and the printer.
- This can slow down image printing and force the use of a hardware image print accelerator card.

www.ArchiveBuilders.com 24002-019

A-171

Production Printing

- Print on demand
 - Manuals, forms
- Customized books
 - Text books, insurance policies
- Direct Mail
- Billing
- Digital blow back
 - Scanned digital images converted back to paper
- Up to 1,300 ppm (*pages per minute*)

www.ArchiveBuilders.com 24002-019

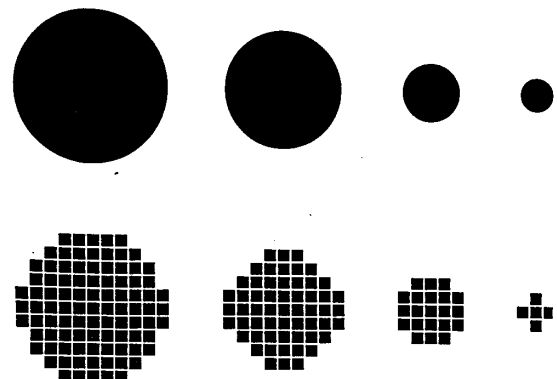
A-172

Fascicular Printing

- Documents Composed and Printed One Page at a Time
- Scitex (www.Scitex.com) MPS42
 - 21 hundred page per minute color inkjet printer
- Xerox (www.Xerox.com) DocuPrint 1300
 - 13 hundred pages per minute Electron beam printer
 - can easily produce over 4 thousand pages per minute
- Xplor International (www.Xplor.org)
 - Professional association for high-speed fascicular printing

www.ArchiveBuilders.com 24002-019

A-173



www.ArchiveBuilders.com 24002-019

Halftone and Laser Dots

A-174

Document Management, Document Imaging

Halftone Printing

- Resolution Given as Screen Ruling
 - 65 - 85 Screen (dpi) for newspapers
 - 133 - 150 Screen for Medium Quality Printing
 - 175 & Up for High Quality Printing
- [A Halftone Dot or Macropel (Macro Picture Element)]
 - Made up of a 10 by 10 to a 16 by 16 grid of Laser Pixels
 - Represents from 101 to 257 theoretical shades of grey
 - Marking Engine Limits Actual Number of Shades of Grey Achieved on a Given Printer
 - [Vernier resolution of 1/2,500th (10 microns)]

www.ArchiveBuilders.com 24002/019

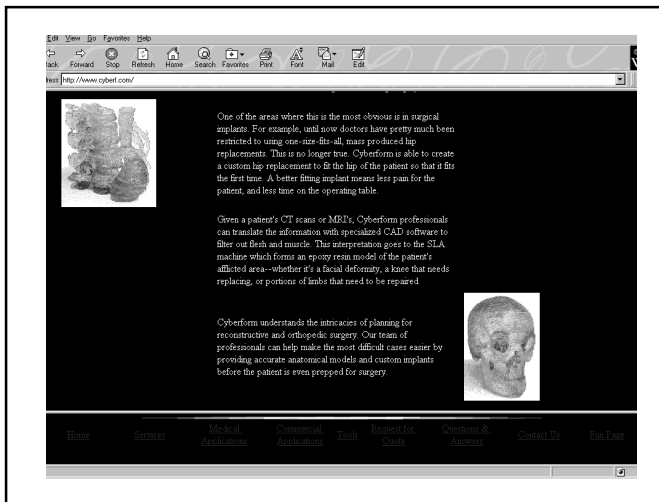
A-175

3D Vectors

- Stereolithography
- Define digital dinosaurs
- Define objects in CAD
 - (C)omputer Aided Design)
- Define objects in GIS (Maps)
 - (G)eographic Information Systems)

www.ArchiveBuilders.com 24002/019

A-176



Stereolithography and 3D (1)

- Convert panoramas and stereo-photos into 3D surfaces
 - VRML(Virtual Reality Modeling Language)
<http://www.VRML.org>
 - e.g. Apple Quicktime VR (Virtual Reality)
<http://developer.apple.com/dev/techsupport/develop/issue25/falco.html>
 - <http://www.apple.com/quicktime/>

www.ArchiveBuilders.com 24002/019

A-178

Stereolithography and 3D (2)

- Printing a 3D copy of the skeleton of a living person to create prostheses
 - <http://www.cyberf.com/>
- Haptic Interface
 - Feel 3D virtual objects
 - Force feedback
 - <http://www.sensable.com/apps.htm>
 - <http://www.sensable.com/apps.htm>

www.ArchiveBuilders.com 24002/019

A-179

Stereolithography and 3D (3)

- Rapid prototyping from CAD
 - <http://206.65.88.5/sla.htm> (www.3DSystems.com)
- Reverse engineering objects
 - by 3D scanning and 3D printing
 - <http://www.millenniumcad.com/home.html>
- Modeling the human genome
 - <http://www.os.kcp.com/home/success/techhum2.html>

www.ArchiveBuilders.com 24002/019

A-180

Document Management, Document Imaging


[Digital Dinosaurs]

- Ray Tracing
- The Frame (Movie or TV Screen)
- The Light
- The Dinosaur
 - Skeleton - Choreography
 - Muscle - Skin
 - Motion - Reflectance

www.ArchiveBuilders.com 24002-019

A-181

Technological Breakthrough

-  Of the 15th century
 - Was as difficult to conceptualize as digitizing
- It was page numbering
- Before page numbering, the end of each printed page carried a snippet of text from the start of the next page
 - This helped binders match up pages when the happy shoppers presented the leaves of their newly purchased books for binding.

www.ArchiveBuilders.com 24002-019

A-182

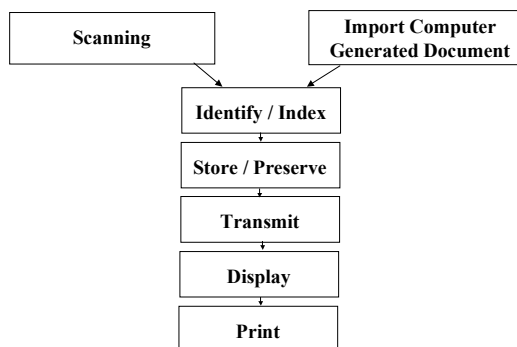
Day/Part Slide Topic

| | | |
|-------|---------|------------------------------------|
| 1 - A | A - 1 | First Principles / Overview |
| 1 - B | A - 59 | Scanning / Printing |
| 1 - C | B - 7 | Importing / Preserving Files |
| 1 - D | B - 186 | Index / Search |
| 2 - A | B - 203 | Store / Purge / Preserve / Display |
| 2 - B | B - 291 | Internet / Systems |
| 2 - C | C - 109 | RFP / Records Management |
| 2 - D | C - 209 | Paradigms, References, & Future |

www.ArchiveBuilders.com 24002-019

A-183

Flow



www.ArchiveBuilders.com 24002-019

A-184

International Association of Aquatic and Marine Science Libraries and Information Centers

IAMSLIC 2000 Conference: Tides of Technology

Workshop: Document Management and Document Imaging for Aquatic and Marine Libraries

September 30 & October 1, 2000, Victoria, British Columbia, Canada

Presented by: Steve Gilheany Tel: +1(310) 937-7000 Fax: +1(310) 937-7001
1209 Manhattan Avenue, C-14, Manhattan Beach, California 90266
<http://www.ArchiveBuilders.com> SteveGilheany@ArchiveBuilders.com

www.ArchiveBuilders.com 24002-019

A-185



IAMSLIC 2000 Conference Tides of Technology

Workshop:

Document Management and Document Imaging for Aquatic and Marine Libraries

September 30 & October 1, 2000, Victoria, British Columbia, Canada

Presented by: SteveGilheany@ArchiveBuilders.com

www.ArchiveBuilders.com 24002-019

A-186

Document Management, Document Imaging



IAMS LIC 2000 Conference Tides of Technology

Workshop: Document Management and Document Imaging for Aquatic and Marine Libraries

September 30 & October 1, 2000, Victoria, British Columbia, Canada

Presented by: SteveGilheany@ArchiveBuilders.com

www.ArchiveBuilders.com 24003-018

B-1

International Association of Aquatic and Marine Science Libraries and Information Centers

IAMS LIC 2000 Conference: Tides of Technology

Workshop: Document Management and Document Imaging for Aquatic and Marine Libraries

September 30 & October 1, 2000, Victoria, British Columbia, Canada

Presented by: Steve Gilheany Tel: +1(310) 937-7000 Fax: +1(310) 937-7001

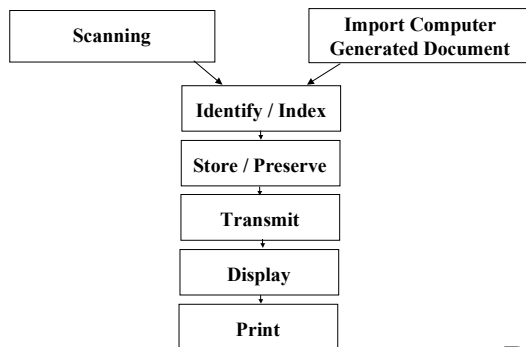
1209 Manhattan Avenue, C-14, Manhattan Beach, California 90266

<http://www.ArchiveBuilders.com> SteveGilheany@ArchiveBuilders.com

www.ArchiveBuilders.com 24003-018

B-2

Flow



www.ArchiveBuilders.com 24003-018

B-3

Daily Schedule

Day 1

- 1 - A
- Break
- 1 - B
- Lunch
- 1 - C
- Break
- 1 - D

Day 2

- 2 - A
- Break
- 2 - B
- Lunch
- 2 - C
- Break
- 2 - D

www.ArchiveBuilders.com 24003-018

B-4

Day/Part Slide Topic

| Day/Part | Slide | Topic |
|----------|---------|------------------------------------|
| 1 - A | A - 1 | First Principles / Overview |
| 1 - B | A - 59 | Scanning / Printing |
| 1 - C | B - 7 | Importing / Preserving Files |
| 1 - D | B - 186 | Index / Search |
| 2 - A | B - 203 | Store / Purge / Preserve / Display |
| 2 - B | B - 291 | Internet / Systems |
| 2 - C | C - 109 | RFP / Records Management |
| 2 - D | C - 209 | Paradigms, References, & Future |

www.ArchiveBuilders.com 24003-018

B-5

Goals of the Session

- To understand format and systems issues in long term electronic records storage, including scanned documents
 - A system designed to be managed
 - Metadata, including indices is required to interpret documents
 - File format integrity: to open a document file, you need the application, the OS (Operating System), and the computer hardware
 - Preserving the bits: media formats are designed to preserve the bits using ECCs (Error Correcting Codes)
- To assist non-technical managers in participating in, and directing, technical discussions.
 - Managers should direct discussions.

www.ArchiveBuilders.com 24003-018

B-6

Document Management, Document Imaging

Importing Computer Generated Files and Their Metadata

www.ArchiveBuilders.com 24003-018

B-7

Metadata

www.ArchiveBuilders.com 24003-018

B-8

And the Score Is . . .

5 to 4

(Metadata)

www.ArchiveBuilders.com 24003-018

B-9

Where Is the Metadata?

- The names of the blank boxes on a form
 - Is the name 'John Jay or 'Jay John?
- The data dictionary for a database
- Field names on a tape record
- The division name on the storage boxes
 - Which division's accounting records are they?
 - (After a merger)
- Finding Aids, Records Management Databases

www.ArchiveBuilders.com 24003-018

B-10

XML, the Language of Metadata

- XML (eXtensible Markup Language) is the language for document metadata.
- The Internet is moving from HTML (HyperText Markup Language) to XML which is a more complete subset of SGML (Structured Generalized Markup Language).

www.ArchiveBuilders.com 24003-018

B-11

What is you Metadata Written In?

- Most metadata is in ASCII.
 - (American Standard Code for Information Interchange)
- Next year Microsoft will leave ASCII completely with Windows 2002
 - Windows NT has always been Unicode based
 - Unicode (16 bits, 65,384 characters)
 - Unicode is a superset of ASCII
 - At <http://www.Unicode.org>

www.ArchiveBuilders.com 24003-018

B-12

Document Management, Document Imaging

Metadata for Metadata about Records

- Metadata is data that explains, for example, which name is the first name and which name is the last name.
- On a punch card, tape, or database flat file: “John Jay” can mean “Jay John” or “John Jay”.
- Using XML (**e**Xtensible **M**arkup **L**anguage) (See Bibliography) metadata tags,
 - “<name:first>John</name:first>” and
 - “<name:last>Jay</name:last>”,
 - the data is unmistakably “John Jay”, now, or at any point in the future.
- John Jay is, in turn, the metadata to access the personnel record for John Jay

www.ArchiveBuilders.com 24003-018

B-13

Metadata

- Comes with the records (documents)
- Stored on fascicles along with the records
 - Written in XML in ASCII (or Unicode)
 - Provides metadata for the metadata
- Fascicles contain metadata for the fascicles
- Fascicles contain metadata for the fascicle TF (**T**rans**F**ormat) system.

www.ArchiveBuilders.com 24003-018

B-14

Chain of Custody

- Written policies and procedures
- System logs (ISO 9000-like audit records)
- Integrated with TF system and its operation
- Based on digital signature maintenance

www.ArchiveBuilders.com 24003-018

B-15

Retention Schedule

- Basis for TransFormat system
- Includes working documents:
 - Records Survey
 - Records Inventory
- Record life-cycles (retention periods) (examples)
 - Permanent (forever) (birth certificate) (land records)
 - Long-term (30 years) (e.g. personnel records)
 - Short-term (7 years) (tax)
 - Working records (1-2 years) (worksheets)
 - Ephemeral records (lunch orders)

www.ArchiveBuilders.com 24003-018

B-16

Website Maintenance

- TF (TransFormat) systems, like all modern software, will provide intranet and perhaps Internet access to records.
- Maintenance of the TF system website that provides this access should be integrated with the metadata maintenance for general records management.
 - For example, the descriptions of records series and their finding aids should be loaded onto the website automatically, from the TF system metadata.
 - If the website information (metadata) is maintained separately from the TF system metadata, then there will be two databases that are supposed to contain the same information.
 - As is always the case, according to Murphy’s Law, if any piece of information is in two databases, it will be different in the two databases.
 - Copying information from one database to another doubles the maintenance burden, at least.
 - All displays of information should be computer retrieved from one master copy of the information.

www.ArchiveBuilders.com 24003-018

B-17

File Format Integrity

(High Level Overview)

www.ArchiveBuilders.com 24003-018

B-18

Document Management, Document Imaging

Grasping at Straws

- The Knee Bone is connected to the Hip Bone, etc.
 - The raster image printed is generated by the printer's RIP
 - RIP (**R**aster **I**mage **P**rocessor)
 - The RIP interprets the PDL page image file
 - (**P**age **D**escription **L**anguage)
 - The document creation application writes the PDL file
 - The document creation application runs on a specific version of the operating system
 - The operating system runs on a specific hardware configuration

www.ArchiveBuilders.com 24003-018

B-19

The Best that Can Be Done

- If Possible, to Preserve a Document, Collect
 - All Related Files
 - Programs: Application and Operating System
 - Emulators
- Moral: Save the Raster Image
 - It is losslessly compressed using a simple, well defined algorithm
 - It can be used to test the rest of the document provenance collected
 - It cannot be reproduced as 'perfectly' as a vector based PDL image
 - It cannot be reproduced as 'perfectly' as a greyscale PDL image

www.ArchiveBuilders.com 24003-018

B-20

Saving the Raster

- What Resolution Raster Image Should be Saved?
 - At Least 600 dpi.
 - RIPped Raster Compresses Very Well Because Computer Generated Images Contain No Digital Noise that is Inserted by Printing, Storing, or Scanning.
- For Want of a Nail, The Horse's Shoe Was Lost . . .
 - Be Sure You Have a Raster Image for Every Page
 - Raster Images are the Bedrock of Low Cost Preservation with Integrity

www.ArchiveBuilders.com 24003-018

B-21

Document as Objects, Composed of Objects (The Richness of Native Formats)

www.ArchiveBuilders.com 24003-018

B-22

Documents Composed of Objects

- Office Suite Document Types
- CAD (**C**omputer **A**ided **D**esign)
- Music, Video, Movies
- GIS (**G**eographic **I**nformation **S**ystems)

www.ArchiveBuilders.com 24003-018

B-23

Document Objects

- Keystrokes,
Mouse Clicks
- Words
- Idioms
- Sentences
- Headings
- Paragraphs
- Subsections
- Chapters
- Sections
- Parts of a Book
- Volumes
- Series
- Languages
- Libraries
- WWW

www.ArchiveBuilders.com 24003-018

B-24

Document Management, Document Imaging

An Object Oriented User Interface

- When you keep clicking in a word processor:
 - First it highlights a character
 - Then it highlights a word
 - Then it highlights a sentence
 - Then it highlights a paragraph
 - Then it highlights a section
 - Then it highlights the entire document

www.ArchiveBuilders.com 24003-018

B-25

Metadata Objects Within Document Files

- Electronic Forms
- SGML Based Documents
(Structured Graphic Markup Language)
 - If you have ever wished a document was formatted consistently in every section
 - SGML is the tool to do it, or to require it.

www.ArchiveBuilders.com 24003-018

B-26

The Lost Facet of Hypertext

- Microhypertext: Characters and Mouse Clicks
 - Are managed as Hyperlinked Objects
- An Infinite Number of Un-do's
- No More 'Whoopses'
 - All Whoopses Can Be Undone
- A Complete History
 - One can Restart Anywhere

www.ArchiveBuilders.com 24003-018

B-27

The 3 Document Formats

www.ArchiveBuilders.com 24003-018

B-28

The 3 Document Format Types

- Native format
 - e.g. Microsoft Word 2000
- Vector format
 - Does not contain data structures to support text and image input.
 - Adobe PDF (**P**ortable **D**ocument **F**ormat) (raster version)
 - Adobe is dedicated to being mathematically correct.
 - (<http://www.Adobe.com>)
- Raster format
 - The perfect raster [not vector] created by a laser printer to render pages of documents sent in vector [not raster] format. [type 2 error]
 - G4 (**G**roup 4) TIFF (**T**agged **I**mage **F**ile **F**ormat)
 - Adobe PDF (**P**ortable **D**ocument **F**ormat) (vector version)
 - Includes raster scanned document images

www.ArchiveBuilders.com 24003-018

B-29

Hierarchy of Format Complexity

- From native format to raster format
 - Each of these formats contains a progressively less rich format, containing progressively less of the original document's structure.
 - Conversely, each of the less rich formats is progressively easier to preserve and present. By storing all of these formats, the maximum benefits of each can be preserved.

www.ArchiveBuilders.com 24003-018

B-30

Document Management, Document Imaging

Lifespans of the 3 Document Formats

- Native format (3 to 9 months)
 - With generous assumptions.
 - No future immutability is offered. None.
 - Not even a millisecond.
- Vector format (30 to 50 years)
- Raster format (500 to 1,000 years)

www.ArchiveBuilders.com 24003-018

B-31

Unadvertised Application Upgrades

- Vendors, such as Microsoft, often do not announce or advertise bug fixes and similar changes
 - The release number often stays the same
 - The only indication of a change is a build number change
 - These changes often affect the appearance of documents when printed or displayed
 - In the first version of Word 2000, printed pagination was different than displayed pagination.
 - Build 9.0.2720

www.ArchiveBuilders.com 24003-018

B-32

Raster Scanned Text -- ('AN')



www.ArchiveBuilders.com 24003-018

B-33

Raster

- A raster format will last forever.
 - Computer hackers view a raster format image as a very lightly-encrypted file.
 - This assures that a raster format file can always be interpreted, as long as the bits are preserved.

www.ArchiveBuilders.com 24003-018

B-34

Native Formats: Will Your Documents Survive the Night?

- Promises from vendors
- Promises from Microsoft
 - <http://msdn.microsoft.com/xml/IE4/License.asp>

www.ArchiveBuilders.com 24003-018

B-35

END-USER LICENSE AGREEMENT FOR MICROSOFT SOFTWARE

IMPORTANT—**READ CAREFULLY**: This Microsoft **End-User License Agreement** ("EULA") is a legal agreement between you (either an individual or a single entity) and Microsoft Corporation for the Microsoft software product identified above, which includes computer software and may include associated media, printed materials, and "online" or electronic documentation ("SOFTWARE PRODUCT"). **By** installing, copying, or otherwise **using the SOFTWARE PRODUCT**, **you agree to be bound** by the terms of this EULA. If you do not agree to the terms of this EULA, do not install or use the SOFTWARE PRODUCT; you may, however, return it to your place of purchase for a **full refund**.

www.ArchiveBuilders.com 24003-018

B-36

Document Management, Document Imaging

Software PRODUCT LICENSE

The SOFTWARE PRODUCT is protected by copyright laws and international copyright treaties, as well as other intellectual property laws and treaties. The SOFTWARE PRODUCT is licensed, **not sold.**

1. GRANT OF LICENSE. This EULA grants you the right to (a) use, modify, and create derivative works of the SOFTWARE PRODUCT in source and object code forms, and (b) reproduce, distribute, sell, rent, or lease copies of the SOFTWARE PRODUCT and derivative works thereof in source and object code forms either as a stand-alone product or included with your own product provided you do not use Microsoft's name, logo, or trademarks to market your product.

2. TERMINATION. Without prejudice to any other rights, Microsoft may terminate this EULA if you fail to comply with the terms and conditions of this EULA. In such event, you must destroy all copies of the SOFTWARE PRODUCT.

3. U.S. GOVERNMENT RESTRICTED RIGHTS. The SOFTWARE PRODUCT and documentation are provided with RESTRICTED RIGHTS. Use, duplication, or disclosure by the Government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.227-7013 or subparagraphs (c)(1) and (2) of the Commercial Computer Software—Restricted Rights at 48 CFR 52.227-19, as applicable. Manufacturer is Microsoft Corporation/One Microsoft Way/Redmond, WA 98052-6399.

www.ArchiveBuilders.com 24003-018

B-37

Software PRODUCT LICENSE

4. EXPORT RESTRICTIONS. You agree that neither you nor your customers intend to or will, directly or indirectly, export or transmit (a) the SOFTWARE PRODUCT or related documentation and technical data, or (b) your Application as described in Section 1 of this EULA (or any part thereof), or process, or service that is the direct product of the SOFTWARE PRODUCT to any country to which such export or transmission is restricted by any applicable U.S. regulation or statute, without the prior written consent, if required, of the Bureau of Export Administration of the U.S. Department of Commerce, or such other governmental entity as may have jurisdiction over such export or transmission.

5. The SOFTWARE PRODUCT is **not fault-tolerant** and is not designed, manufactured, or intended for use or resale as on-line control equipment in hazardous environments requiring fail-safe performance, such as in the operation of nuclear facilities, aircraft navigation or communication systems, air traffic control, direct life support machines, or weapons systems, in which the failure of the SOFTWARE PRODUCT could lead directly to death, personal injury, or severe physical or environmental damage.

www.ArchiveBuilders.com 24003-018

B-38

MISCELLANEOUS

If you acquired this product in the United States, this EULA is governed by the laws of the State of Washington.

If you acquired this product in Canada, this EULA is governed by the laws of the Province of Ontario, Canada. Each of the parties hereto irrevocably attorns to the jurisdiction of the courts of the Province of Ontario and further agrees to commence any litigation which may arise hereunder in the courts located in the Judicial District of York, Province of Ontario.

If this product was acquired outside the United States, then local law may apply.

Should you have any questions concerning this EULA, or if you desire to contact Microsoft for any reason, please contact the Microsoft subsidiary serving your country, or write: Microsoft Sales Information Center/One Microsoft Way/Redmond, WA 98052-6399.

www.ArchiveBuilders.com 24003-018

B-39

LIMITED WARRANTY

NO WARRANTIES. Microsoft expressly disclaims any warranty for the SOFTWARE PRODUCT. The SOFTWARE PRODUCT and any related documentation is provided "**as is**" **without warranty of any kind**, either express or implied, including, without limitation, the implied warranties or merchantability, **fitness for a particular purpose**, or **noninfringement**. The **entire risk** arising out of use or performance of the SOFTWARE PRODUCT **remains with you**.

www.ArchiveBuilders.com 24003-018

B-40

LIMITED WARRANTY (Cont.)

NO LIABILITY FOR DAMAGES. In no event shall Microsoft or its suppliers be liable for any damages whatsoever (including, without limitation, damages for loss of business profits, business interruption, **loss of business information**, or any other pecuniary loss) arising out of the use of or inability to use this Microsoft product, even if Microsoft has been advised of the possibility of such damages. Because some states/jurisdictions do not allow the exclusion or limitation of liability for consequential or incidental damages, the above limitation may not apply to you.

I have read and **approved** the end user license agreement.

www.ArchiveBuilders.com 24003-018

B-41

Protection

- The courts believe that agreements in writing provide protection.
- The courts believe that the ability to abrogate, provided in writing, provides protection.
 - Market share is important.
 - Just another paradigm.

www.ArchiveBuilders.com 24003-018

B-42

Document Management, Document Imaging

Dangers of the Native Format (1)

- When software application version numbers change:
- (And sometimes when it does not change):
 - Work-arounds are now the inverse (opposite) of the problem they fixed.
 - Fonts, letter spacing, special symbols, line spacing, and page layout change imperceptibly.
 - Except that the last word on some lines does not fit:
 - Creating new lines and changing the pagination:
 - Changing the Table of Contents and the Index: maybe?, hopefully?

www.ArchiveBuilders.com 24003-018

B-43

Dangers of the Native Format (2)

- When version numbers change:
 - Graphics disappear
 - Words and concepts are no longer grouped by boxes
 - Graphics appear when bugs are fixed
 - Special text positioning entries become redundant and confusing.

| |
|--------------|
| California |
| New York |
| Delaware |
| Rhode Island |

becomes

California
New York
Delaware
Rhode Island

www.ArchiveBuilders.com 24003-018

B-44

Dangers of the Native Format (3)

- Some native formats do not necessarily include a viewable image
 - **CAD (Computer Aided Design)**
 - CAD is a 3 dimensional model.
 - 2 dimensional projections (drawings) must be specified.
 - Specification is impossible if the native application cannot be run.
 - **GIS (Geographic Information System)**
 - A GIS is a 3 dimensional digital terrain model.
 - A GIS can include the entire earth.
 - 2 dimensional projections (maps) must be specified.
 - Specification is impossible if the native application cannot be run.
 - **CAD and GIS files are almost never linked:**
 - Even though CAD and GIS files are very similar.
 - Therefore it is still easy to build structures across lot boundaries.
 - CAD documents can be linked to a GIS by using 3 or more control points to geo-reference the CAD models to the GIS.

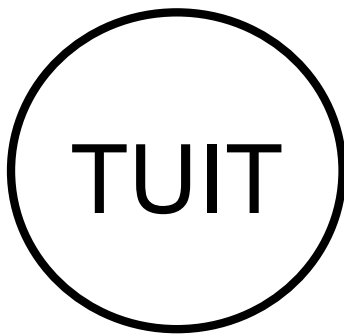
www.ArchiveBuilders.com 24003-018

B-45

TUIT

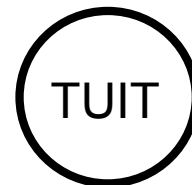
www.ArchiveBuilders.com 24003-018

B-46



www.ArchiveBuilders.com 24003-018

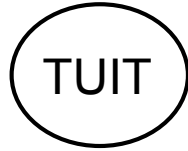
B-47



www.ArchiveBuilders.com 24003-018

B-48

Document Management, Document Imaging



www.ArchiveBuilders.com 24003-018

B-49

The Disappearing 'TUIT' Circle

- The circle is projected correctly in PowerPoint 97.
- The circle disappears when printed for an overhead.
- The circle is round on an 8 1/2 x 11 size, 2 or 6 up handout.
- The circle is an ellipse on an A4, metric size, 2 or 6 up handout.
- The next version of PowerPoint may correct one or more of these effects. (Revisions are not always publicly identified.)
- Subsequent version may correct different combinations of these effects.
- Preserving a raster bit map of each format solves the problem.

www.ArchiveBuilders.com 24003-018

B-50

Why Preserve the Native Format If it Has So Many Problems?

- Required for low-cost republishing
- Required for low-cost updates
- Hidden changes can be avoided because:
 - The author will proof for glitches
 - A comparison of the native format raster output and the vector and raster format raster outputs will provide warnings of where changes have occurred.

www.ArchiveBuilders.com 24003-018

B-51

Republishing / Re-Engineering

- A professional author is required.
- Liability is based on changes in intellectual content.
- The native file format is required.
- Any change requires that the professional author also review the entire document for changes caused by a new native application revision number.
 - Records managers have a professional responsibility to advise authors of this responsibility.
- Republishing / Re-Engineering is often used to recover (recreate) paper files for which the electronic copy has been lost.

www.ArchiveBuilders.com 24003-018

B-52

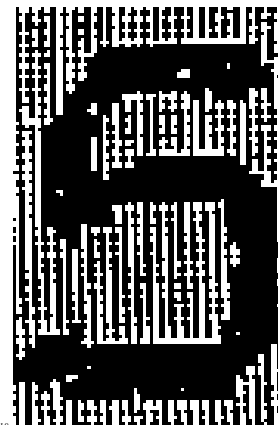
Muddling Through

- Compare the original computer generated image with the new computer generated raster image that came from the new native application or emulating viewer
 - Fix the layout in the native application
 - Because you cannot change the viewer or the new version of the native application
- Muddling through supports:
 - Re-authoring
 - Re-engineering
 - Re-researching, re-developing

www.ArchiveBuilders.com 24003-018

B-53

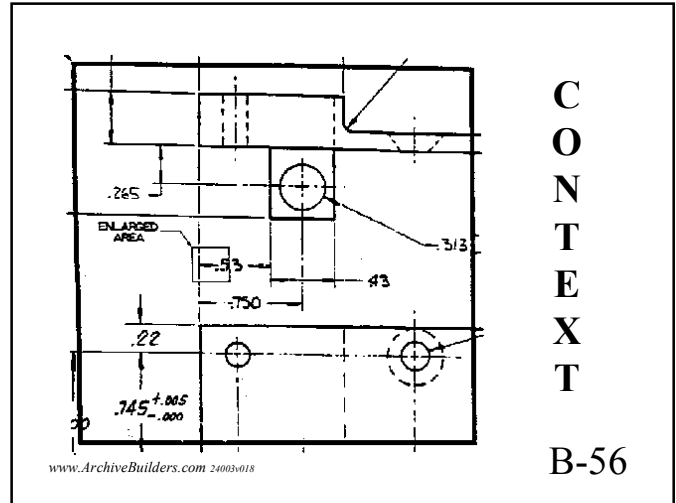
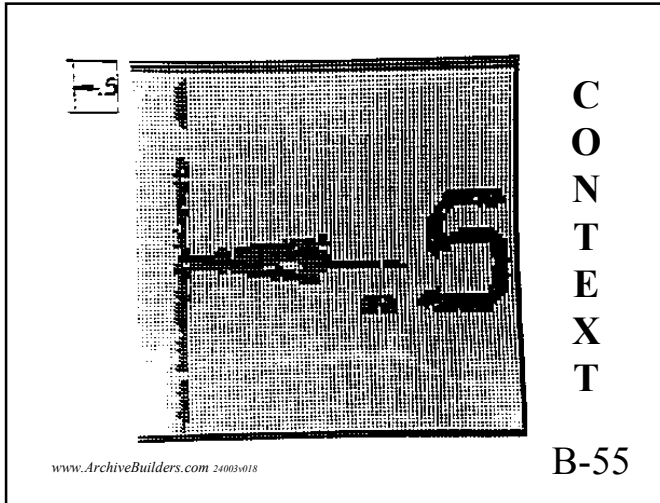
What
is it?



www.ArchiveBuilders.com 24003-018

B-54

Document Management, Document Imaging



Printing: Outline Fonts

- When you print a scanned raster image, you just reproduce the scanned pixels.
- When you print a word processor file, the printer generates the pixels from your text.
 - For each character, the location, font, and size is sent to the printer.
 - The printer draws an outline of each character.
 - Outline Fonts
 - The printer then fills in the outline with pixels and prints the character.

www.ArchiveBuilders.com 24003-018

B-57

Outline Fonts

- The Size, Font, and Location of Character Glyphs on a Page is Conveyed to a Printer in a PDL File
 - PDL (Page Description Language, e.g. Postscript)
- Characters are Generated as Vector Outlines
 - Character Outlines are Made Up of Segments that are Defined by Third Order Equations. The Outline Segments are Butted so that the Butt is Smooth (the Second Derivative is Zero).
 - The Vector Processing is Computationally Floating Point Intensive
- The Vector Outlines are Filled in With Pixels
 - The Fill Pixels are Registered to the Printer Pixel Grid
 - The Vector Outlines carry 'hints' to Match the Characteristics of the Printer Marking Engine

www.ArchiveBuilders.com 24003-018

B-58

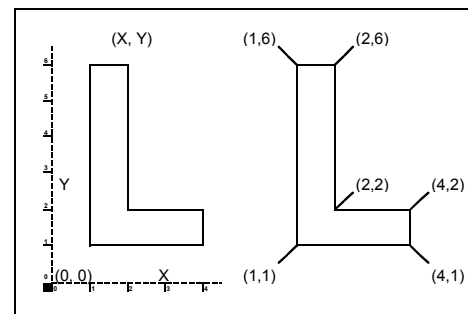
Vectors

- Connect point 'A' to point 'B'
- Horizontal Vectors
- Vertical Vectors
- Diagonal Vectors
- Circles
- Curvy Line Vectors
- Vectors form the outline of all visible graphics

www.ArchiveBuilders.com 24003-018

B-59

An 'L' produced by the following outline vector list coordinates: {(1,6); (2, 6); (2, 2); (4,2); (4,1); (1, 1)}

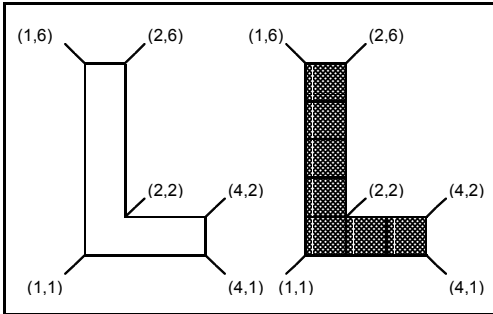


www.ArchiveBuilders.com 24003-018

B-60

Document Management, Document Imaging

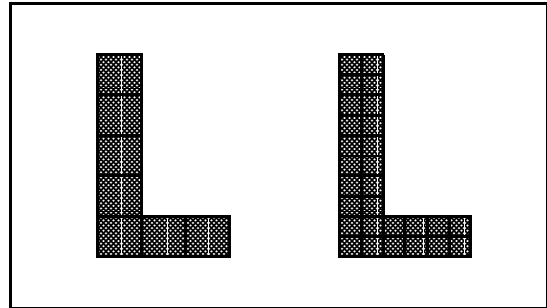
Filling-in an 'L' outline font with large pixels, creates a computer generated 'L' character glyph



www.ArchiveBuilders.com 24003-018

B-61

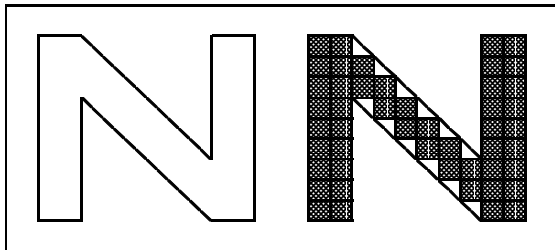
Filling in a second 'L' with smaller pixels



www.ArchiveBuilders.com 24003-018

B-62

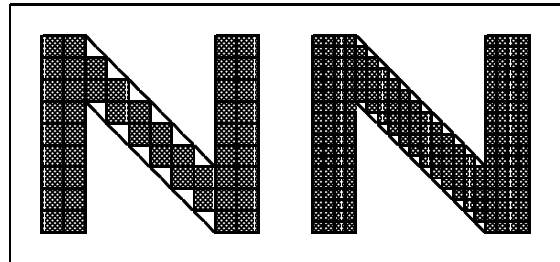
An 'N' glyph filled with large pixels



www.ArchiveBuilders.com 24003-018

B-63

A computer generated 'N' character glyph filled-in with smaller pixels



www.ArchiveBuilders.com 24003-018

B-64

Mini Pics: 'Lil Vehicles' Outline Font

a b c d e f g h
Arial Outline Font

a b c d e f g h
Times Roman Outline Font



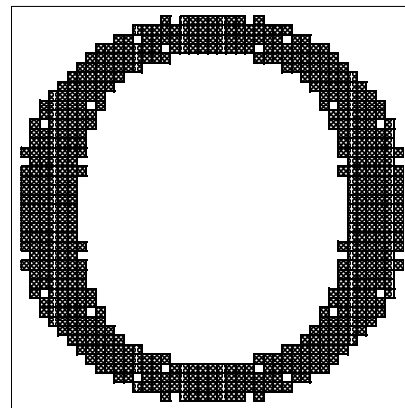
Mini Pics: Lil Vehicles Outline Font



Raster Image

www.ArchiveBuilders.com 24003-018

B-65



www.ArchiveBuilders.com 24003-018

Crafted Edge Character: 'O'

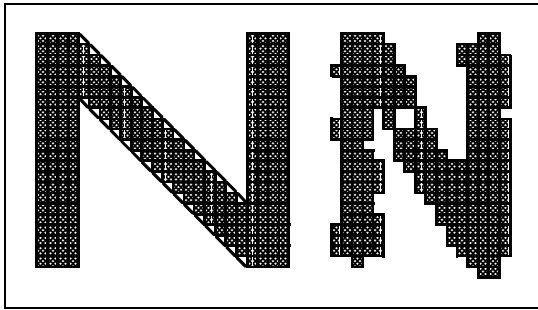
- the inside circle of the character 'O' of radius 'R';
- $X^2 + Y^2 = R^2$ forms a circle of radius R
- The outside of the 'O' with a stroke thickness of 'T': $(X+T)^2 + (Y+T)^2 = (R+T)^2$
- But most typeset 'O's have a stroke of varying thickness which requires X cubed and Y cubed terms in their equations.

14 Point ITC Korinna Regular Uppercase 'O' in 300 dpi

B-66

Document Management, Document Imaging

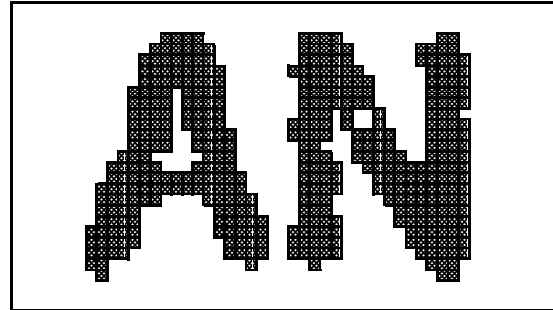
Computer generated glyph of an 'N'
beside a raster scanned glyph of an 'N'



www.ArchiveBuilders.com 24003-018

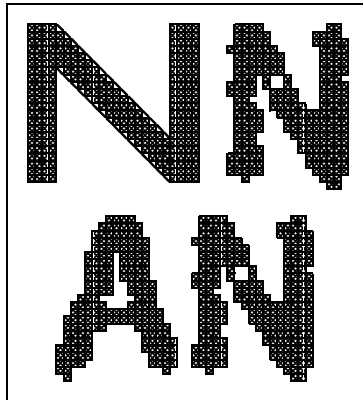
B-67

Raster scanned glyph of an 'A' placed next to
raster scanned glyph of an 'N' of the same type font



www.ArchiveBuilders.com 24003-018

B-68

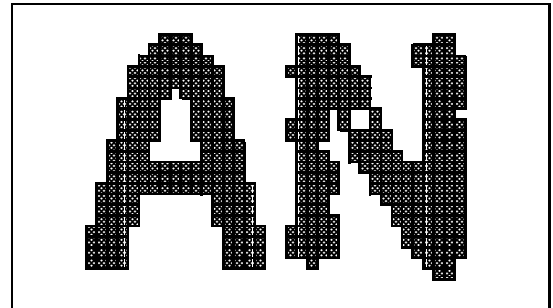


www.ArchiveBuilders.com 24003-018

Computer
generated
glyph (N)
and scanned
glyph (A),
both beside
a raster
scanned
glyph (N)

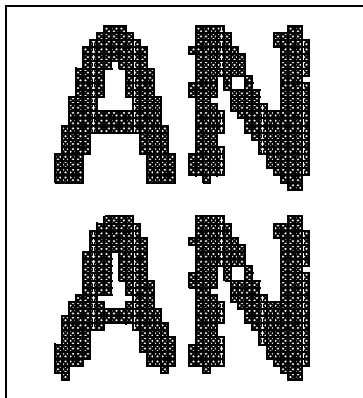
B-69

A computer generated 'A' glyph, that exactly matches
the font that the 'N' glyph was scanned from



www.ArchiveBuilders.com 24003-018

B-70



www.ArchiveBuilders.com 24003-018

The two 'N'
glyphs are
identical, but
the top 'N'
glyph, next to
the computer
generated 'A'
glyph, looks
much better.

B-71

Computer Generated Images

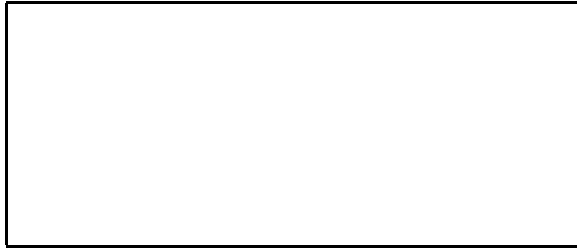
- This computing power is also used to generate the images of text and graphics for computer generated documents such as:
 - Word processor documents
 - Spreadsheets and slide presentations
 - CAD (Computer Aided Design)
 - Database reports
 - GIS (Geographic Information System)
 - Gant Charts and PERT charts for project scheduling
 - PERT (Performance Evaluation Review Team)
 - Created by Admiral Hyman G. Rickover for the US Nuclear Submarine program in the early 1950s.

www.ArchiveBuilders.com 24003-018

B-72

Document Management, Document Imaging

Illustration of Document Digitizing

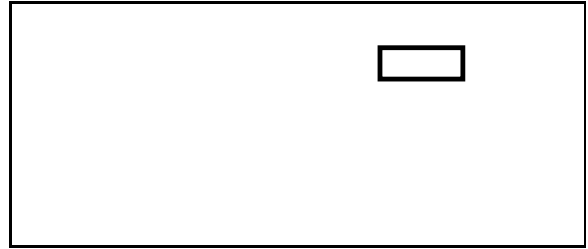


From white paper 22021, *How Digitizing Works* at the website below

www.ArchiveBuilders.com 24003-018

B-79

Illustration of Document Digitizing

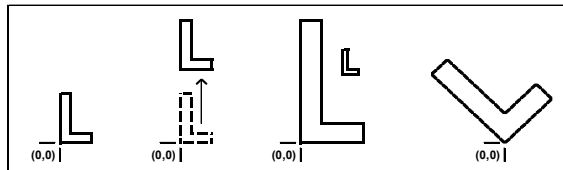


From white paper 22021, *How Digitizing Works* at the website below

www.ArchiveBuilders.com 24003-018

B-80

Movement, Scaling, and Rotation (Matrix)



$$\begin{bmatrix} a & b & 0 \\ c & d & 0 \\ t & u & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t & u & 1 \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\text{Theta}) & \sin(\text{Theta}) & 0 \\ -\sin(\text{Theta}) & \cos(\text{Theta}) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

General
Matrix

Movement
Matrix

Scaling
Matrix

Rotation
Matrix

www.ArchiveBuilders.com 24003-018

B-81

General Transformation

- This represents a simultaneous movement, scaling, and rotation.
- The coordinate pair (x, y) can be transformed into another coordinate pair (x', y') by multiplying the (x, y) coordinate pair by the general transform matrix.

$$\text{➤ } x' = ax + cy + t$$

$$\text{➤ } y' = bx + dy + u$$

www.ArchiveBuilders.com 24003-018

B-82

Movement (Spatial Translation)

- When the movement matrix is used, a spatial translation or movement (of 't' in the 'x' dimension and 'u' in the 'y' dimension) is effected.
- The movement is 't' units in the 'x' direction.
- The movement is 'u' units in the 'y' direction.
- Movement
 - $x' = x + t$
 - $y' = y + u$

www.ArchiveBuilders.com 24003-018

B-83

Scaling

- The scaling matrix produces scaling (enlargement or reduction of the image size)
- By a factor of 'a' in the 'x' dimension.
- By a factor of 'd' in the 'y' dimension.
- Scaling
 - $x' = xa$
 - $y' = yd$

www.ArchiveBuilders.com 24003-018

B-84

Document Management, Document Imaging

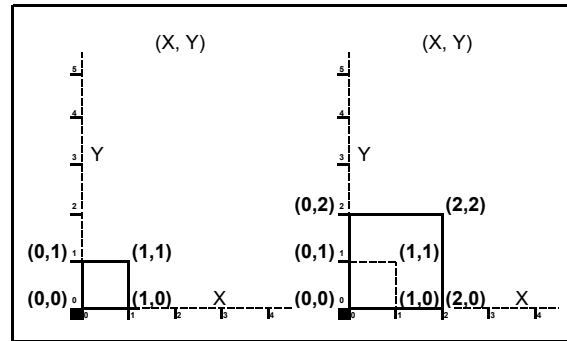
Rotation

- Use of the rotation matrix produces counterclockwise rotation of the image about the origin by an angle Theta (θ).
 - $x' = \cos(\theta) x + \sin(\theta) y$
 - $y' = -\sin(\theta) x + \cos(\theta) y$

www.ArchiveBuilders.com 24003-018

B-85

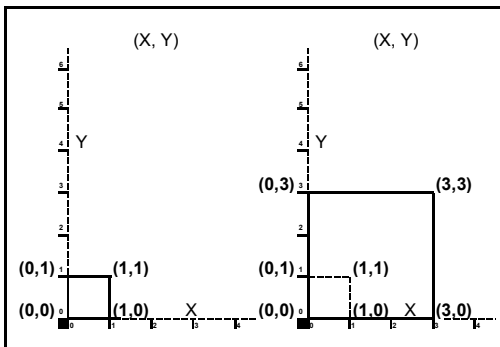
Scaling up of a symbol for a period



Polygon $\{(0,1), (1,1), (1,0), (0,0)\}$ is multiplied by $(2,2)$, doubling its size to $\{(0,2), (2,2), (2,0), (0,0)\}$

B-86

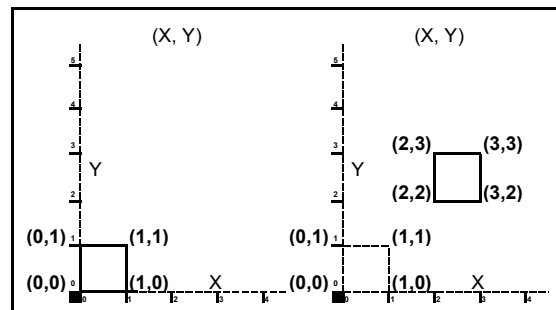
Scaling up of a symbol for a period



Polygon $\{(0,1), (1,1), (1,0), (0,0)\}$ is multiplied by $(3,3)$, tripling its size to $\{(0,3), (3,3), (3,0), (0,0)\}$

B-87

Moving a symbol for a period up and to the right



By adding $(2, 2)$ to the polygon $\{(0,1), (1,1), (1,0), (0,0)\}$ it is moved to $\{(2,3), (3,3), (3,2), (2,2)\}$

B-88

Applying Everything at Once

- Rotate
 - Turn
- Scale
 - Shrink
- Translate
 - Move
- Repeat (Copy)
 - A child playing with a rubber stamp
- All together
 - Spiral-in

www.ArchiveBuilders.com 24003-018

B-89



Copyright 1981, by Scott Kim, all rights reserved. In Inversions, ISBN 1-55953-280-7 www.ScottKim.com Used with permission

B-90

Document Management, Document Imaging

Spiral Infinity, Exemplar for PostScript

- **Spiral Infinity** is the exemplar of the Interpress (Xerox), of the PostScript (Adobe), and later of the Acrobat PDF (Portable Document Format) (Adobe) outline font based PDLs (Page Description Languages).
- The Spiral Infinity outline font image was created by Scott Kim, <http://www.ScottKim.com>, and John Warnock, <http://www.Adobe.com>, in 1981, before John Warnock founded Adobe Systems, in 1982.
- This image illustrates the mathematical nature of the outline character fonts used to reproduce almost all computer generated textual documents today.

www.ArchiveBuilders.com 24003-018

B-91

Spiral Infinity, Exemplar for PostScript (cont.)

- The spiral effect can only be created by manipulating the underlying mathematical formulas that define the outlines of the type in this image.
- This image was created to show everything that PostScript would become, using a single image, before the language was written, and to illustrate the mathematics of, and the now pervasive nature (after 20 years) of, PDL based typography, graphics, and graphics editing.

www.ArchiveBuilders.com 24003-018

B-92

ASCII Image and ASCII Metadata

- Most metadata is in ASCII. (8 bit bytes)
 - (American Standard Code for Information Interchange)
- The ‘Spiral Infinity’ and ‘Illustration of Document Digitizing’ images are both in ASCII
 - The graphics files are written in PDF
 - All PDF is in ASCII
- Microsoft is moving the Unicode. (16 bit bytes)

www.ArchiveBuilders.com 24003-018

B-93

(From a Bit and Pixel Oriented Perspective) Raster File Formats: Group 4 Fax, TIFF, and PDF

- TIFF is a header on a group 4 fax file
- PDF is a header on a group 4 fax file
- Both TIFF and PDF use the non-destructive group 4 fax compression and raster image format.
 - The group 4 fax compression format is what hackers can recognize.

www.ArchiveBuilders.com 24003-018

B-94

TIFF

www.ArchiveBuilders.com 24003-018

B-95

TIFF: a Families of Formats

- TIFF (*Tagged Image File Format*)
 - Is a family of formats defined by a standard tag in a file header
 - The tag defines the rest of the header
 - Some TIFF formats can only be read by proprietary readers that erase the file after one reading
 - There is a TIFF Bible to assist in interpreting the hundreds of TIFF formats
 - Adobe now assigns TIFF tag numbers
 - TIFF was created by Aldus Inc. in the 1980s and Aldus was subsequently purchased by Adobe on August 31, 1996
 - The Library of Congress assigns Dewey Decimal Numbers

www.ArchiveBuilders.com 24003-018

B-96

Document Management, Document Imaging

Adobe PDF File Format

www.ArchiveBuilders.com 24003-018

B-97

PDF: a Families of Formats

- PDF (Adobe's **P**ortable **D**ocument **F**ormat)
 - Is a family of formats defined by Adobe
 - New formats can be added to the family at any time
 - Only Adobe can add new PDF formats
 - Adobe owns the copyright to the word 'PDF'
- 2 of the PDF formats
 - Text (Vector based outline fonts)
 - Raster (Hidden text)
 - Hides OCR (**O**ptical **C**haracter **R**ecognition) text behind the raster image
 - Uses the OCR text for full text search and highlight hits on the raster image

www.ArchiveBuilders.com 24003-018

B-98

Five PDF Formats

1. Raster scanned image, no OCR produced text
2. Raster scanned image, OCR text hidden, stored as hidden text
3. Raster scanned image, OCR text replaces raster glyphs in Normal format
4. Computer generated (outline font) raster image
5. Outline fonts in a vector format

www.ArchiveBuilders.com 24003-018

B-99

Why PDF over TIFF (1)

- TIFF (*Tagged Image File Format*)
 - The TIFF standard allows, and has encouraged, the creation of thousands of document file formats.
 - Open Standard
 - Free for All
 - Many competitive hooks have been put in by vendors to make document formats proprietary.
 - .TIF is the 3 letter file extension.

www.ArchiveBuilders.com 24003-018

B-100

Why PDF over TIFF (2)

- PDF (**P**ortable **D**ocument **F**ormat)
 - Closed Proprietary Format
 - Good
 - Like Windows
 - Adobe will sue anyone creating nonstandard PDF files.
 - Adobe is devoted to mathematical precision and consistency.

www.ArchiveBuilders.com 24003-018

B-101

Why PDF over TIFF (3)

- Simple
 - Easy to understand and specify
 - Like Windows
- Caution
 - Be sure to avoid the 'normal' format.
 - Be sure to specify 'hidden text'.
 - Be sure your systems people keep the flag set to 'hidden text'.

www.ArchiveBuilders.com 24003-018

B-102

Document Management, Document Imaging

The Wonders of PDF (1)

- PDF (**P**ortable **D**ocument **F**ormat) from Adobe
- 'Normal' format destroys the integrity of the document image.
 - The 'normal' format is only useful for republishing or re-engineering.
 - Records Management does not do republishing or reengineering.
- 'Hidden text' is the only format setting for managing liability.
 - The user sees the raw scanned image.
 - The user must use their professional judgement to evaluate the image and its potential for error and liability.

www.ArchiveBuilders.com 24003-018

B-103

The Wonders of PDF (2)

- PDF (**P**ortable **D**ocument **F**ormat) from Adobe
 - Vector format is derived from Postscript
 - Print to PDF
 - Vector format
 - Full text searchable with no errors
 - Simple edit are possible.
 - Most edits require use of native file format, not PDF.

www.ArchiveBuilders.com 24003-018

B-104

Adobe Acrobat

- Creates PDF (**P**ortable **D**ocument **F**ormat)
- PDF format is derived from the Postscript language
- Can mix raster and vector characters
 - Masks OCR errors
 - Causes users to overlook their responsibility for interpretation
- Acrobat has the best registration for OCR'd text.
 - The best for placing hot links
 - The best for locating full text search hits

www.ArchiveBuilders.com 24003-018

B-105

Adobe Acrobat Input Formats

- Print Image (PDL - **P**age **D**efinition **L**anguage)
 - Adobe Postscript (Output from CAD, GIS, office applications)
 - Microsoft TrueType
 - HP (Hewlett Packard) PCL5 (**P**rinter **C**ontrol **L**anguage)
- Scanned Raster
 - Raster plus separate OCR page image (Hidden Text)
 - Raster mixed with OCR characters (Normal)
- Printer or screen bitmap (not scaleable)
 - Only available by importing PDF text file into Adobe PhotoShop

www.ArchiveBuilders.com 24003-018

B-106

Document Formats for CAD

- Native format (**C**AD - **C**omputer **A**ided **D**esign)
 - 3D (Dimensional) Model (not inherently viewable)
 - DWG (AutoCAD from Autodesk)
 - DGN (MicroStation from Bentley/Intergraph)
 - 2D Printed Format (definition of a specific view)
 - DWF (AutoCAD)
 - .SVF (MicroStation)
- Vector format
 - Adobe PDF (**P**ortable **D**ocument **F**ormat)
- Raster format
 - Adobe PDF (**P**ortable **D**ocument **F**ormat)
- An electronic signature on a drawing is good for only as long as the format of the drawing file can be read.

www.ArchiveBuilders.com 24003-018

B-107

ESRI

- About 70 percent of the world GIS market
 - (**G**eographic **I**nformation **S**ystem)
- Tens of billions of US dollars are spent creating maps with ESRI software each year.
- **E**nvironmental **S**ystems **R**esearch **I**nstitute, Inc.
- <http://www.ESRI.com>
- Pronounced 'es-ree'
- Products: ArcInfo, ArcView
- Headquartered in Redlands, CA

www.ArchiveBuilders.com 24003-018

B-108

Document Management, Document Imaging

A History of Emulators: Intel

- Each Intel chip emulates all previous Intel chips.
- Chips and operating systems are developed using emulators before the new, target chip is ready.
 - 4004 (4bit), 4040, 8008 (8 bit), 8080, 8086 (16 bit) 80186, 80286, 80386 (32 bit), 80486, 80586 (Pentium), 80686 (Pentium II), 80786 (Merced) (64 bit)
- Bill Gates developed his Basic interpreter on a software emulator of an Intel 8080 that ran on a DEC (Digital Equipment Corporation) PDP 10 computer.

www.ArchiveBuilders.com 24003-018

B-109

Ptolemaic Epicycles (1)

- All means of expression are being recreated in digital form.
- All techniques of reproducing expression are being recreated in digital form.
- The complexity of digital records is growing along all known dimensions.

www.ArchiveBuilders.com 24003-018

B-110

Ptolemaic Epicycles (2)

- The good news is:
 - The growth is somewhat bounded.
- Current growth in complexity is mostly a product of converting existing techniques to digital form.
 - When this is completed, only new forms and techniques will have to be handled.
 - These forms and techniques will likely appear at their historic, much slower rate.

www.ArchiveBuilders.com 24003-018

B-111

Rosetta Stone

- “The text of the stone included instructions that the document should be set up in hard stone in the three scripts which the Rosetta Stone still bears today (hieroglyphic, Demotic, and Greek)”. (March, 196 BC)
 - Thus, the Rosetta Stone contains its own metadata, and a single document reproduced in three formats that are locked (sealed) together in stone.
 - The three formats were intended to help preserve the mystique of the, then, 3 thousand year hieroglyphic format, which would only last (be understood) another 4 hundred years.
- British Museum, London, collection number EA 24
- Also used to connote a digital medium that carries a raster image of the text to be conveyed to the future. (Norsam.com)

www.ArchiveBuilders.com 24003-018

B-112

Storage Estimates

www.ArchiveBuilders.com 24003-018

B-113

Where Are Your Electronic Records?

- On which disk?
- On which server?
- On which SAN? (Storage Area Network)
- On which backup?
 - Is it an incremental backup?
 - Is it a full backup?
 - Is it a volume backup?
 - Is its retention schedule compliant?
- How is it protected from:
 - Access (retrieval, alteration)?
 - Destruction (denial of service)?

www.ArchiveBuilders.com 24003-018

B-114

Document Management, Document Imaging

Storage Estimates

- These estimates will help you focus your meetings on design issues rather than whether a page is 50 or 51 KiloBytes.
- After you have scanned less than 1 percent of your images you will have a very exact estimate of your storage requirements.
 - If your images are 10 per cent larger, the you will need 10 percent more storage, all of you other design decision will probably stay the same.

www.ArchiveBuilders.com 24003-018

B-115

Imaging Facts

- **1 scanned page (8 1/2 by 11 inches, A4) = 50 KiloBytes**
 - (on average, black & white, CCITT G4 compressed)
- 1 file cabinet (4 drawer) (10,000 pages on average) = 500 MegaBytes (MByte) = 1 CD (Compact Disc) (ROM or WORM)
- 1 box (in inches: 15 1/2 long x 12 wide x 10 deep) (400 x 300 x 250 mm) (2,500 pages) = 1 file drawer = 125 MegaBytes
- 1 box (packed) = 2 linear feet (500 mm) of files (loose enough for active filing) = 25 (rounded) linear inches = 125 MegaBytes
- 1 linear inch (~20 mm) = 100 pages = 5 MegaBytes
- 1 cubic foot (cf) (~.025 cubic meter) = 2000 pages = 100 MegaBytes

www.ArchiveBuilders.com 24003-018

B-116

Color Photos

- 1 color picture = 10 KiloBytes (thumbnail)
- 5 MegaBytes
 - (for each of 100 photos on a 500 MegaByte Kodak PhotoCD)

www.ArchiveBuilders.com 24003-018

B-117

Scanned Microforms

- 1 roll of 16 mm microfilm (100 ft, ~30 meters) (24X reduction) = 2,500 letter size images = 1 box = 1 file cabinet drawer = 125 MegaBytes
- 1 roll of 35 mm microfilm (100 ft) (12X reduction, open spacing, normal scan) = 1,000 letter size images = 50 MegaBytes
- 1 microfiche (105 mm film) (24X reduction) = 100 letter size images = 5 MegaBytes (average); 200 microfiche = 20,000 images = 1 GigaByte

www.ArchiveBuilders.com 24003-018

B-118

One Fiche or Roll per Record

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

www.ArchiveBuilders.com 24003-018

B-119

Pixels per Image

- Scanned aperture card images require the same storage as images scanned from the paper original of the document in the aperture.
 - This is true for any microform
- No matter how much a photograph is enlarged or reduced, the storage required is determined by the quality of the camera lens and film used to make the original photograph, microphotograph (microform), or photomicrograph (blowback).
 - Image quality may be reduced by subsequent changes in image size, but image quality can never be increased.

www.ArchiveBuilders.com 24003-018

B-120

Document Management, Document Imaging

Office Color

- 1 scanned page = 100 KiloBytes
 - (100 dpi) (8 1/2 by 11 inches, A4)
 - On average, office color, including grayscale, compressed
- View-only-color (100 dpi) (no OCR possible)
 - About twice the size of document images scanned in a bi-tonal, black and white format, and then G4 compressed.
- Has subsumed grayscale scanning
 - Office color document file formats may include a 300 dpi bi-tonal, black and white scan of the same image.
- No standards yet

www.ArchiveBuilders.com 24003-018

B-121

Video

- 2 hour video on one layer of 1 side of a DVD = 4 GigaBytes
- 1 hour of DVD quality video = 2 GigaBytes

www.ArchiveBuilders.com 24003-018

B-122

Audio

- 1 hour of audio = 10 MegaBytes
 - (dictation, answering machine)
- 1 hour of audio = 500 Mbytes
 - (a 650 MegaByte CD holds 74 minutes of music)

www.ArchiveBuilders.com 24003-018

B-123

Color Depth Makes Documents More Readable

- Office quality color scanning
 - Scanned color differences aid users in reading a document
 - Increasing the quality of OCR (Optical Character Recognition) done at 150 dpi and higher resolutions
 - If 3D (Dimensional) (Grayscale) OCR is used.

www.ArchiveBuilders.com 24003-018

B-124

Chest X-Ray

- 1 Chest X-ray = 1 MegaByte
 - (14 x 17 inches)
 - 150 dpi (dots per inch), 12 bits (compressed)
 - Wavlet compression, lossless mode, has FDA 510(k) approval.
 - (12 bits per pixel provide 4,096 shades of gray.)
 - (150 dpi, 12 bit images are recommended by the American College of Radiology for primary reads.)

www.ArchiveBuilders.com 24003-018

B-125

Aerial Photography Digital Orthophotography (1)

- 500 square mile city using 6 inch pixels
 - 50 GigaPels (50 billion pixels)
 - 4 pixels per square foot
 - 100 million pixels per square mile
 - 25 million square feet per square mile
- Using 24 bit color and estimating a lossless three-to-one compression, this digital orthophotographic image would require 50 GigaBytes to store.

www.ArchiveBuilders.com 24003-018

B-126

Document Management, Document Imaging

Aerial Photography Digital Orthophotography (2)

- 500 square mile city using 2 inch pixels
 - 9 times as many pixels
 - 450 GigaPels (450 billion pixels)
- 50 square mile city using 2 inch pixels
 - 45 GigaPels requiring (45 GigaBytes)
 - Color would require 45 GigaBytes to store
 - using the same compression assumptions.

www.ArchiveBuilders.com 24003018

B-127

GIS

- **Geographic Information System**
- 50 MegaBytes of data per square mile
 - 500 square mile city is about 25 GigaBytes
 - Excluding Digital Orthophotographs

www.ArchiveBuilders.com 24003018

B-128

Satellite and Aerial Photography And Remote Sensing

- Weather photographs, crop quality assessment, or storm damage / flooding
- 24 bit color or multispectral images
- Land area (size): 1 thousand kilometers by 1 thousand kilometers (600 miles by 600 miles)
- Uses 100 meter pixels (400 foot pixels)
 - (pixels that are 100 meters by 100 meters)
 - (400 feet by 400 feet)
- Contains 100 million pixels
 - 100 MegaBytes uncompressed grayscale image
 - 100 MegaBytes compressed color image

www.ArchiveBuilders.com 24003018

B-129

Non-Scanned Pages

- All pages other than wordprocessor pages, OCRed pages, and COLD pages are scanned pages.

www.ArchiveBuilders.com 24003018

B-130

OCR and Wordprocessor Pages

- 1 wordprocessor page:
 - = 1 OCRed page
 - (**O**ptical **C**haracter **R**ecognition)
 - = 5 Kbytes

www.ArchiveBuilders.com 24003018

B-131

COLD Pages

- 1 COLD page
 - (Computer **O**utput to **L**aser **D**isc)
 - Compressed
 - Includes index
 - Also includes COOL (Computer **O**utput **O**n-**L**ine)
- = 2 KiloBytes for letter size statements
- = 4 KiloBytes for 11 x 14 inch fanfolded greenbar computer sheet
- = 10 KiloBytes for **A**ll **P**oints **A**ddressable (**APA**) pages such as IBM AFP (**A**dvanced **F**unction **P**rinting) and Xerox Metacode.

www.ArchiveBuilders.com 24003018

B-132

Document Management, Document Imaging

PDF Pages

- Adobe PDF (Portable Document Format)
 - 1 Page
 - = 10 KiloBytes for pages that are all text
 - = 50 KiloBytes for graphics
 - = 1 MegaByte for images
 - Please note that these are general estimates
 - A small sample will improve the estimate greatly
 - However, these general estimates work well for most design requirements.

www.ArchiveBuilders.com 24003-018

B-133

Unicode

- The foundation for unified, single, worldwide software releases.
 - Microsoft's interest
- Unicode bytes are:
 - 16 Bits
 - Allows for 65,536 Different Characters
 - Supports most of the world's languages
- A superset of ASCII
 - (American Standard Code for Information Interchange)
- At <http://www.Unicode.org>

www.ArchiveBuilders.com 24003-018

B-134

Converting Trees to Records

- 1 pulp tree (loblolly pine)
 - = 1/10th cord of wood
 - = 10,000 pages
 - = 1 File Cabinet
 - = 4 boxes
 - = 1/2 GigaByte
 - = 1 CD

www.ArchiveBuilders.com 24003-018

B-135

Imported Computer Generated Documents

www.ArchiveBuilders.com 24003-018

B-136

Computer Generated Documents

- Word Processor
- Spreadsheets
- Slide shows
- Comma delimited flat file from database
- Maps
- Blueprints, engineering drawings

www.ArchiveBuilders.com 24003-018

B-137

Preservation Forever

- Permanent retention
 - Getting the records back in the box (fascicle)
 - Where the records can be protected and managed
 - Where are your electronic records right now?
- Electronic records deteriorate over time
 - Bits fade away (like disintegrating acid paper)
 - Formats fade away (like lost languages)
 - Systems and hardware fade away (like knotted documents)
 - Knowledge and integrity face away (chain of custody)
- Fascicles
 - Electronically signed virtual digital containers

www.ArchiveBuilders.com 24003-018

B-138

Document Management, Document Imaging

Importing and Preserving Computer Files

www.ArchiveBuilders.com 24003-018

B-139

Imported Computer Files

- Computer Generated Files
- Fax
- Computer Generated Files With Raster
- COLD
- EDI (Electronic Data Interchange)
- Adobe Acrobat (.PDF)
- Email / RichText / DVD Formats

www.ArchiveBuilders.com 24003-018

B-140

Document Structure

- Partial recovery from scanned documents
- Reconstruction occurs at many levels
- Correcting errors
 - Not in scanned theses
- Fixing up a somewhat structured document
- Complete creation of structure
 - From the ground up
 - With the intent to 'look like' the original

www.ArchiveBuilders.com 24003-018

B-141

Recovering Structure (1)

- Never Perfect
- Frequently worse than not doing it
- Must be considered a reauthoring of the document.
- Frequently the only acceptable solution is to completely restructure and even rewrite the entire document.

www.ArchiveBuilders.com 24003-018

B-142

Recovering Structure (2)

- You may even create a structure for a group of documents.
 - Then you must convert each document to match the structure.

www.ArchiveBuilders.com 24003-018

B-143

Sequence of Recovery

- Raster Image
- OCR, OCR correction
- SGML
- Every document element may have its own structural definition
- No two document elements may be related
- Fix the worst problems

www.ArchiveBuilders.com 24003-018

B-144

Document Management, Document Imaging

CAD Sequence of Recovery

- Raster Image
- OCR, OCR correction
- Vectorize, Complete Vectors
- Snap vectors to grid
- Match written dimensions to drawing
 - Or vice versa
- Match up 2D views, fix discrepancies
- Project to 3D
- Create Solid Model

www.ArchiveBuilders.com 24003-018

B-145

CAD Recovery

- Even less rewarding than recovering text documents.
- The most common solution is to vectorize just the portion of a raster image being updated.
 - The person modifying the drawing can sign-off on the correctness of the vectorization.
 - If the new vectors are converted back to raster, the intelligence of the modification is lost.

www.ArchiveBuilders.com 24003-018

B-146

Sequence of Printing

- Ignore native editing hooks (such as ‘undo’ push-down stacks, and comment authorizations)
- Use text tags to flow text into SGML structure
- Establish location (including negative leading), graphic (e.g.: character, font, font size, font morphing), graphic morphing, color, and page morphing
- Map image to pixel raster of printer
- Print

www.ArchiveBuilders.com 24003-018

B-147

Recovering Databases

- Database recovery is beyond the scope of this presentation in most cases
- Databases can be converted to comma delimited flat files and stored as documents in all the formats listed above (especially raster).
 - This will allow the data to persist for a much longer time than if the data was locked in a procedural database.

www.ArchiveBuilders.com 24003-018

B-148

Importing Documents as Files & Preservation

www.ArchiveBuilders.com 24003-018

B-149

2 Preservation Issues

- Intellectual (Interpreting complex formats)
 - Covered here, under file formats
 - If you had the bits, could you interpret them?
- Mechanical (Protecting the bits)
 - Covered later in the presentation, under storage
 - Making sure that you do not lose any bits
 - Based on longevity of media and error correcting codes

www.ArchiveBuilders.com 24003-018

B-150

Document Management, Document Imaging

Computer Generated Files

- Word Processor
- Spreadsheet
- Office Suite
- CAD (Computer Aided Design)
- CAM (Computer Aided Manufacturing)
- GIS (Geographic Information Systems)

www.ArchiveBuilders.com 24003-018

B-151

Fax

- Fax Server
- DID (**D**irect **I**nward **D**ialing)
 - Everyone gets their own ghost fax phone number
- CSID (**C**all **S**ubscriber **I**dentification)
 - Blocked with call blocking
- Fax Routing / Filing

www.ArchiveBuilders.com 24003-018

B-152

Computer Generated Files with Raster

- POD (**P**roof of **D**elivery)
- Electronic Signatures
- Signatures as Characters

www.ArchiveBuilders.com 24003-018

B-153

COLD

- Makes Mainframe Reports PC Accessible
 - COLD (**C**omputer **O**utput to **L**aser (optical) **D**isk)
 - COOL (**C**omputer **O**utput **O**n-**L**ine)
 - COM (**C**omputer **O**utput **M**icrofilm)
- Integrated with Document Imaging
- Reverse Engineer the Database
 - Based on Printed Reports
- Full Text Search of Printed Reports

www.ArchiveBuilders.com 24003-018

B-154

CAD

- Many Buildings are Designed to Last 40 Years
 - CAD (**C**omputer **A**ided **D**esign)
- Aperture Cards Strip Out Objects and Hyperlinks
- Hyperlinks Require Locatable Objects
- Retention Schedules
- We Adhere to All Applicable Regulations
- OS Versions to Support CAD Versions
 - OS (**O**perating **S**ystem of a computer)
- CALS (**C**omputer **A**ided **L**ogistics)
 - US DoD (**D**epartment of **D**efense)

www.ArchiveBuilders.com 24003-018

B-155

Document Formats

- Native format (Microsoft Word 97) (Many Issues)
- Vector format
 - Adobe Postscript
 - Adobe PDF (Portable Document Format)
- Raster format
 - G4 (**G**roup 4) TIFF (**T**agged **I**mage **F**ile **F**ormat)
 - Adobe PDF (**P**ortable **D**ocument **F**ormat)

www.ArchiveBuilders.com 24003-018

B-156

Document Management, Document Imaging

[National GIS Program]

- Video of a National GIS program
- CAD provides a framework to hold previously scattered or non-existent data elements.

www.ArchiveBuilders.com 24003-018

B-157

[GIS and CAD for User Orientation]

- Video / Demonstration
 - <http://www.Multigen-Paradigm.com>
- Images on buildings and walls
- Virtual museum
- Spatial Indexing
- Traveling through time
 - 4D (four *d*imensional) simulacrum: p46-58, in Arthur C. Clark's, *The City and the Stars*, Harcourt, Brace & World, 1956.

www.ArchiveBuilders.com 24003-018

B-158

GIS: Geographic Information Systems

www.ArchiveBuilders.com 24003-018

B-159

Geo-Reference CAD to GIS

- Link (Geo-Reference) CAD and GIS
 - The house on the hill is really on the hill
 - And — walls do not cross property boundaries
 - Interferences are detected automatically
- Facilitates Virtual Reality (VR)
 - Fly-overs (and fly-unders)
 - Fly-forward (and fly-back) in time
 - Animations of change

www.ArchiveBuilders.com 24003-018

B-160

Linking Paper Documents

- Old maps and drawings
 - appear as 2D projections floating in the 3D CAD and GIS space
 - A road profile blueprint appears as a plane cutting through the road.
- Textual documents are indexed by the parcel(s) that they describe
 - They appear as a hot spots on a layer of the GIS
 - Or as voxels (volume elements) in an aquifer
 - Or as flashing 3D rooms in a burning building

www.ArchiveBuilders.com 24003-018

B-161

Stereo Photo ➡ Panorama ➡ 3D

- Apple's QuickTime VR (*V*irtual *R*eality)
- Aerial photography now has GPS (*G*lobal *P*ositioning *S*ystem) control points in the plane.

www.ArchiveBuilders.com 24003-018

B-162

Document Management, Document Imaging

Stereo Photo ➡ Panorama ➡ 3D

- Hand held cameras will have the same capabilities in 10 to 20 years
 - You move the camera around, the camera snaps pictures when it is pointed in the right direction
 - Creating a 3D digital image (surface)
 - Solid state 'inertial' navigation will allow interior shots away from GPS line of sight satellite transmission.

www.ArchiveBuilders.com 24003-018

B-163

As Built Records

- Stereo cameras create a 3D digital image.
- These 3D images create 3D as-built records for:
 - Buildings
 - Substructures (hazardous pipelines and power lines)
 - Earthworks

www.ArchiveBuilders.com 24003-018

B-164

As Built Records

- Use the computer to add or subtract these 3D as-built images from the planned CAD and GIS 3D images
 - The variances show up as red volumes
 - on the CAD and GIS screen

www.ArchiveBuilders.com 24003-018

B-165

Data Field ➡ Push Down Stack

- Every database field becomes a push down stack
 - Changes are pushed onto the stack
 - Changes are time stamped
 - One can check the value of a field at a given time
 - One can check the state of an entire database at a given time
 - Documenting fixes becomes more complicated

www.ArchiveBuilders.com 24003-018

B-166

Environmental Impact

- All information presented electronically
 - No more paper
 - More trees survive the planning process
 - Punch cards did disappear, after years of proof that they would never disappear. (And now we have forgotten.)
- The Internet replaces delivery of paper
 - No more trips to the planning office
 - Disabled persons can work from home
 - Or — even the hospital or rest home

www.ArchiveBuilders.com 24003-018

B-167

ESRI

- About 70 percent of the world GIS market
- Tens of billions of dollars (USD) are spent creating maps with ESRI software each year.
- Environmental Systems Research Institute, Inc.
- <http://www.ESRI.com>
- Pronounced 'es-ree'
- Products: ArcInfo, ArcView
- Headquartered in Redlands, CA

www.ArchiveBuilders.com 24003-018

B-168

Document Management, Document Imaging

ADA **(Americans with Disabilities Act)**

- Your plans are ADA compliant
- Is your documentation ADA compliant?
 - Does your document system shift colors into the color space of colorblind persons, automatically
 - Can your document be printed in 3D using stereolithography for unsighted persons
 - Computers can already do it
 - It is only necessary to ask

www.ArchiveBuilders.com 24003-018

B-169

Unicode Bytes

are

16 Bits

(65,536 Different Characters)

(Are They Accessible?)

www.ArchiveBuilders.com 24003-018

B-170

Unicode

- Unicode 3.0 just completed.
- <http://www.Unicode.org/>

www.ArchiveBuilders.com 24003-018

B-171

Vector

- Carries the complexity of handling all recording formats
 - Can be made tractable by using emulation to recreate pixels and vectors
 - Postscript Level 3 and SGML provide insights into this complexity

www.ArchiveBuilders.com 24003-018

B-172

Levels of Electronic Preservation

- A function of an Archive, not Records Management
- Preserving all the steps in document printing.
 - The raster image that is printed is generated by the RIP.
 - The RIP interprets the PDL page image.
 - RIP (*R*aster *I*mage *P*rocessor) PDL (*P*age *D*escription *L*anguage)
 - The document creation application writes the PDL file.
 - The document application runs on a specific version of the operating system (e.g. Microsoft Windows 2000).
 - The operating system runs on a specific hardware configuration.

www.ArchiveBuilders.com 24003-018

B-173

Digital Libraries and Archives

www.ArchiveBuilders.com 24003-018

B-174

Document Management, Document Imaging

Libraries, Archives & Museums

- Libraries
 - Preserve Opinions (Descriptions, Histories, and Analysis)
- Archives
 - Preserve Source Materials, Unprocessed
 - For future researchers recording their considered opinions. (Descriptions, Histories, and Analyses)
- Museums
 - Museums are the archives of physical objects.
 - Museums provide interpretation and analysis of their collected objects.

www.ArchiveBuilders.com 24003-018

B-175

Libraries, Archives, and Museums

- They all seek to preserve.
- They all seek to provide access.
- Libraries emphasize access.
- Archives emphasize preservation.
- Museums emphasize physical objects.
- All three provide a means of going back to the original record of past events and reinterpreting the events.
 - History does not exist independently. There is no 'is' – or 'proof'.
 - History is an interpretation that only exists within a paradigm.
- The goals of all three begin to merge in the digital world.

www.ArchiveBuilders.com 24003-018

B-176

Archives

- Preserving Accountability
- Sealed files can avoid copyright issues
 - A proven archives technique applied to the digital age.
- Developing digital media to meet archival needs.
- Professional responsibility to know the technology.
 - And how to employ technology to carry out professional responsibilities.

www.ArchiveBuilders.com 24003-018

B-177

Should You Convert to a Common Format?

- Limit conversion, avoid migration
 - TF (TransFormat) system administrators should copy, not migrate
 - TF system users (not TF system administrators) should do all format conversions, and submit documents in all suggested formats.
 - The user is responsible to ensure that the digital document of record is also the legal document of record, which is in use by the TF system user's organization
- Common Formats
 - Raster (G4 - Group 4 fax)
 - HTML (HyperText Markup Language)
 - SGML (Structured Generalized Markup Language)
 - Book
 - Libraries encourage their production
 - Libraries do not create books

www.ArchiveBuilders.com 24003-018

B-178

Migratory Effects: Are there (Insurmountable) Problems?

- Just ask Mr. Twelve-Foot
 - The migrated Mark Twain
 - Technology *W*ithout an Interesting Name
 - TWAIN is a set of drivers for slow slow scanners. <http://www.TWAIN.org/>
 - ISIS (*I*mage and *S*canner *I*nterface *S*pecification) is a set of drivers for fast scanners. <http://www.PixTran.com/>
 - The NBI word processors were *N*othing *B*ut *I*nitials.
- It is said that the Archbishop of Seville believed that all of science and technology was in the etymology of language.
 - To *orient* a map, place the east at the top.
 - The La Brea Tar Pits – in Spanish or English

www.ArchiveBuilders.com 24003-018

B-179

Professional Responsibilities of Librarians, Archivists, and Museum Curators

- Fairness
- Balance
- Integrity

www.ArchiveBuilders.com 24003-018

B-180

Document Management, Document Imaging

Choices

- Choices must be made
 - Libraries
 - Collection Development
 - Archives
 - Accessioning

www.ArchiveBuilders.com 24003-018

B-181

The Third Millennium

www.ArchiveBuilders.com 24003-018

B-182

Materials Science & Mechanics



Abiomed

www.ArchiveBuilders.com 24003-018

B-183

Permanent Artificial Heart

- Ready for clinical trials soon
- In the US, 60 thousand people per year need a heart transplant
 - 2 thousand heart transplants are done
- By Abiomed
 - A Massachusetts company

www.ArchiveBuilders.com 24003-018

B-184

Words

- The Second Millennium was the last millennium for computer things 'Mega'
 - MegaBytes and MegaHertz are going the way of KiloBytes and KiloHertz
 - Now there will be only GigaBytes of RAM and GigaHertz processor speeds
 - GigaByte was an unknown three years ago, now TeraByte is appearing.
 - TeraByte = 1 thousand GigaBytes

www.ArchiveBuilders.com 24003-018

B-185

Index / Search

www.ArchiveBuilders.com 24003-018

B-186

Document Management, Document Imaging

Document Identification

- Done by scanner operator
- Not the same as document indexing
- Minimum information necessary to link paper document with digital document images
- Also done by Bates numbers
 - Bates stamp marks each page with a sequential number and automatically advances for next page
 - Pages can be marked before or after scanning

www.ArchiveBuilders.com 24003-018

B-187

Search

- There are many search methods
- Some use existing indices
- Some require extensive intellectual effort
- Some are fully automated
- As the field matures all techniques are becoming equally available and useable
 - Combining techniques is very powerful

www.ArchiveBuilders.com 24003-018

B-188

Access Via a Key

- Where to start?
 - Paper files that all have the same key, such as a customer number, are the low hanging fruit.
 - Start with them.

www.ArchiveBuilders.com 24003-018

B-189

Access Via Databases Structures

- Database Structures:
 - Flat File
 - Hierarchical
 - Relational
 - Object Oriented
- Image Enabled Applications

www.ArchiveBuilders.com 24003-018

B-190

Access Via File Room Metaphor

- File Room Metaphor:
 - File Room
 - File Cabinet
 - File Drawer
 - File Folder
 - Nested File Folder
 - Document

www.ArchiveBuilders.com 24003-018

B-191

Access Via an Alias or Shortcut

- An alias or a shortcut is a virtual document that is actually a pointer to the real document.

www.ArchiveBuilders.com 24003-018

B-192

Document Management, Document Imaging

Access Via an Alias or Shortcut

- An alias or a shortcut is a virtual document that is actually a pointer to the real document.
- Avoids storing the document file in more than one folder for the purpose of entering the document in two or more folders.

www.ArchiveBuilders.com 24003-018

B-193

Document Indexing (1)

- Full Text
- Fuzzy Search
- Nested Folders
- Sequential
- Date Scanned
- Key Word
- Unique Identifier (Key)
- Email Link
- Workflow Link
- Log of Reading History
- Active-X Object
- Bibliography
- Citation Counts
- Database Entry

www.ArchiveBuilders.com 24003-018

B-194

Document Indexing (2)

- Aliased/Shortcut Folder
- Card Catalog
- Finding Aid
- Time Code (SMPTE)
- Internet Agent
- Concept/Thesaurus
- Structure (SGML)
- Hyperlink Traversal
- Hyperlink Annotation
- Popularity Chart
- Genealogy
- GIS/GPS Coordinates
- CAD Object Reference
- Combination of Above

www.ArchiveBuilders.com 24003-018

B-195

Retrieval Terms

- Precision (% of hits that are relevant.)
- Recall (% of available hits retrieved.)
- Narrow the Search (Reduce the number of hits by being more specific.)
- Broaden the Search (Increase the number of hits by relaxing constraints.)

www.ArchiveBuilders.com 24003-018

B-196

Indexing Tools and Techniques

- Full Text
 - Fall-back in case of RIF (Layoff) (Reduction in Force)
- Thesaurus
- Affinity (Terms that commonly appear together.)
- Authority Lists (Specific Instances, Names)
- Controlled Vocabulary (Generic)
- Back of Book Indexing (Customized)

www.ArchiveBuilders.com 24003-018

B-197

Advantage of Full Text Over Custom Indexing

- Oil companies originally indexed their documents to facilitate their search for oil.
 - These indices were of little use when pollution control became of paramount importance.
- Full text indexing does not discriminate based on search criteria.
- Combining the techniques is best.

www.ArchiveBuilders.com 24003-018

B-198

Document Management, Document Imaging

Sequential Numbering is Best (1)

- For accessioning and organizing there is nothing better than sequential numbering.
 - Most files are organized so that expensive detailed indexing is replaced by sequential searches, shifting expenses to the time of the search and to the search budget.
- Sequential numbers can be added to digital records without physically changing the records.

www.ArchiveBuilders.com 24003-018

B-199

Sequential Numbering is Best (2)

- In dictionaries, we go to about the right place and then search sequentially.
- For documents organized in legal cases sequential Bates numbers are assigned to each page.
 - Bates was the manufacturer of the auto-incrementing manual page numbering stamp.

www.ArchiveBuilders.com 24003-018

B-200

Hypertext

- Hypertext hyperlinks can be made and annotated at every point in the multidimensional definition space of documents.

www.ArchiveBuilders.com 24003-018

B-201

Spatial Indexing

- Index novels by their location
- Show all movie clips for a given filming location.

www.ArchiveBuilders.com 24003-018

B-202

Store / Purge

www.ArchiveBuilders.com 24003-018

B-203

Pixels in an Inch & Square Inch

- 1 pixel per inch produces 1 pixel per square inch
- 2 pixels per inch produces 4 pixels per square inch
- 3 pixels per inch produces 9 pixels per square inch
- 10 pixels per inch produces 100 pixels per square inch
- 200 pixels per inch produces 40,000 pixels per sq. in.
- 300 pixels per inch produces 90,000 pixels per sq. in.
- 400 pixels per inch produces 160,000 pixels per sq. in.
- 600 pixels per inch produces 360,000 pixels per sq. in.

www.ArchiveBuilders.com 24003-018

B-204

Document Management, Document Imaging

United States Paper Sizes in Inches, New (Trimmed) and Old (Untrimmed)

| Name | New Size | Old Size | Name | New Size | Old Size |
|------|------------|----------|------|---------------------|----------|
| A | 8 1/2 x 11 | 9 x 12 | F | 28 x 40 | varies |
| B | 11 x 17 | 12 x 18 | G | 11 x (22 1/2 to 90) | varies |
| C | 17 x 22 | 18 x 24 | H | 28 x (44 to 143) | varies |
| D | 22 x 34 | 24 x 36 | J | 34 x (55 to 176) | varies |
| E | 34 x 44 | 36 x 48 | K | 40 x (55 to 143) | varies |

www.ArchiveBuilders.com 24003-018

B-205

Metric Trimmed Paper Sizes

| Metric Name | Metric Size in mm | Size in Inches | United States Size | Size in Inches |
|-------------|-------------------|----------------|--------------------|----------------|
| A8 | 53 x 74 | 2.07 x 2.91 | Business card | 2 x 3 1/2 |
| A7 | 74 x 105 | 2.91 x 4.13 | 3 x 5 | 3 x 5 |
| A6 | 105 x 148 | 4.13 x 5.83 | Microfiche | 4.13 x 5.83 |
| A5 | 148 x 210 | 5.83 x 8.27 | 5 x 8 | 5 x 8 |
| A4 | 210 x 297 | 8.27 x 11.69 | A | 8 1/2 x 11 |
| A3 | 297 x 420 | 11.69 x 16.54 | B | 11 x 17 |
| A2 | 420 x 594 | 16.54 x 23.39 | C | 17 x 22 |
| A1 | 594 x 841 | 23.39 x 33.11 | D | 22 x 34 |
| A0 | 841 x 1189 | 33.11 x 46.81 | E | 34 x 44 |
| 2A | 1189 x 1681 | 46.81 x 66.22 | | |

www.ArchiveBuilders.com 24003-018

B-206

Boxes to GigaBytes

- 1 Page = 50 KiloBytes when stored
– (Document images are always stored and transmitted in compressed form.)
- 1 File Cabinet (4 Drawer) = 10,000 Pages
- 1 File Cabinet = 1 CD ROM (1/2 GByte)
- 1 Box = 2,500 pages
- 8 Boxes = 2 File Cabinets = 1 GigaByte
- 8,000 Boxes = 2,000 File Cabinets = 1 TeraByte

www.ArchiveBuilders.com 24003-018

B-207

Cubic Feet to GigaBytes

- 10 cubic feet = 1 GigaByte = 8 boxes
- 5 cubic feet = 1 file cabinet = 4 boxes
- 10 thousand cubic feet = 1 TeraByte

www.ArchiveBuilders.com 24003-018

B-208

Linear Feet to GigaBytes

- 8 linear feet = 1 file cabinet = 4 boxes
- 16 linear feet = 1 GigaByte = 8 boxes
- 16 thousand linear feet = 1 TeraByte

www.ArchiveBuilders.com 24003-018

B-209

... Bytes

- 1 KiloByte = 1,000 Bytes = 1 Thousand Bytes
- 1 MegaByte = 1,000 KiloBytes = 1 Million Bytes
- 1 GigaByte = 1,000 MegaBytes = 1 Billion Bytes
- 1 TeraByte = 1,000 GigaBytes = 1 Trillion Bytes
- 1 PetaByte = 1,000 TeraBytes = 1 Quadrillion Bytes
- 1 ExaByte = 1,000 PetaBytes = 1 Quintillion Bytes
- 1 ZettaByte = 1,000 ExaBytes = 1 Sextillion Bytes
- 1 YottaByte = 1,000 ZettaBytes = 1 Septillion Bytes
 - Due to lawsuits, only the ISO metric 1,000 multiples are used in commercial presentations.

www.ArchiveBuilders.com 24003-018

B-210

Storage Costs

www.ArchiveBuilders.com 24003/018

B-211

Magnetic Disk Storage Cost Over the Next 10 Years

| Annual Decline | Cost For 1 GigaByte (US Dollars) (Storage for 2 Scanned File Cabinets) | Cost For 1 TeraByte = 1,000 GigaBytes (US Dollars) (Storage for 2,000 Scanned File Cabinets) (Holding 20 Million Scanned Letter Size Pages) | | | | |
|----------------|--|---|--|-----------------------------|---|---------------|
| | | Non-FC/SCSI PC Disk No Online Redundancy | Non-FC/SCSI PC Disk Software RAID Redundancy | SAN FC Fabric Hardware RAID | SCSI/FC SAN/PC Name Brand Fault Awareness Hardware RAID | Mainframe |
| Year | File Cabinets | 1 X | 2 X | 4 X | 8 X | 12 X |
| 1992 | 1,000.00 | 1,000,000.00 | 2,000,000.00 | 4,000,000.00 | 8,000,000.00 | 12,000,000.00 |
| 1993 | 550.00 | 550,000.00 | 1,100,000.00 | 2,200,000.00 | 4,400,000.00 | 6,600,000.00 |
| 1994 | 302.50 | 302,500.00 | 605,000.00 | 1,210,000.00 | 2,420,000.00 | 3,630,000.00 |
| 1995 | 166.38 | 166,375.00 | 332,750.00 | 665,500.00 | 1,331,000.00 | 1,996,500.00 |
| 1996 | 91.51 | 91,506.25 | 183,012.50 | 366,025.00 | 732,050.00 | 1,098,075.00 |
| 1997 | 50.33 | 50,328.44 | 100,656.88 | 201,313.75 | 402,627.50 | 603,941.25 |
| 1998 | 27.68 | 27,680.64 | 55,361.28 | 110,722.56 | 221,445.13 | 332,167.69 |
| 1999 | 15.22 | 15,224.35 | 30,448.70 | 60,897.41 | 121,794.82 | 182,692.23 |
| 2000 | 8.37 | 8,373.39 | 16,746.79 | 33,493.58 | 66,987.15 | 100,480.73 |
| 2001 | 4.61 | 4,605.37 | 9,210.73 | 18,421.47 | 36,842.93 | 55,264.40 |
| 2002 | 2.53 | 2,532.95 | 5,065.90 | 10,131.81 | 20,263.61 | 30,395.42 |
| 2003 | 1.38 | 1,393.12 | 2,786.25 | 5,572.49 | 11,144.99 | 16,717.48 |
| 2004 | 0.77 | 766.22 | 1,532.44 | 3,064.87 | 6,129.74 | 9,194.61 |
| 2005 | 0.42 | 421.42 | 842.84 | 1,685.68 | 3,371.36 | 5,057.04 |
| 2006 | 0.23 | 231.78 | 463.56 | 927.12 | 1,854.25 | 2,781.37 |
| 2007 | 0.13 | 127.48 | 254.96 | 509.92 | 1,019.84 | 1,529.75 |
| 2008 | 0.07 | 70.11 | 140.23 | 280.45 | 560.91 | 841.36 |
| 2009 | 0.04 | 38.56 | 77.13 | 154.25 | 308.50 | 462.75 |
| 2010 | 0.02 | 21.21 | 42.42 | 84.84 | 169.68 | 254.51 |

www.ArchiveBuilders.com 24003/018

B-212

Store

- Magnetic Disk
- Optical Disk
- Security / Integrity
- System Support
- Scalability

www.ArchiveBuilders.com 24003/018

B-213

Magnetic Disk

www.ArchiveBuilders.com 24003/018

B-214

Magnetic Disk

- Spelled 'disk', with a 'k'
 - Following the metal 'disk' in a harrow
- RAID (Redundant Array of Inexpensive Disks)
 - Mirroring (Copy)
 - Parity disk (To reconstruct a damaged disk)
 - Hot swap (pluggable) of a drive sled into a drive holster
 - Striping (for speed)
 - A good example of an ECC (Error Correcting Code)
- FailOver (To a twin processor in a cluster)
- Replication of databases at remote sites

www.ArchiveBuilders.com 24003/018

B-215

Fiber Channel Disk Interface

- Gigabit per second interface
- Very small connector
 - Like new telephone jack
- Over one hundred disks on one controller card
- Lengths in kilometers (miles)
 - SAN (System / Storage Area Network)

www.ArchiveBuilders.com 24003/018

B-216

Document Management, Document Imaging

Fiber Channel Fabric

- Like having the entire Internet in a computer center
- Up to 16 million disks per fabric
 - On a single PC connection
 - 1 ExaBytes = 1 million TeraBytes
 - Using 73 GigaByte disks

www.ArchiveBuilders.com 24003-018

B-217

8 TeraBytes of RAID

- US\$ 300 Thousand (rack & disks, disks purchased 6/99 to 6/00)
- 120 73 GigaByte disks (out about June 1999, www.Seagate.com)
 - 10 thousand RPM (**R**evolutions **p**er **M**inute)
 - 6 millisecond access, 36 MegaBytes per second, per drive, transfer rate
- 1 Fiber Channel connection for 120 drives
 - 100 MegaBytes/second (1 pages/second to each of 2,000 users)
 - up to 10 kilometers (5 miles) between drives
 - Global hot spares for all RAID arrays on the controller
- 1 19 inch rack (500 mm) (900 pounds, 400 kilograms, 6 sq. ft., 1 sq. meter)
 - Like movable shelves, a full 19 inch rack exceed floor loading limits.
 - 6 ft. tall (2 meters)
- 1 TeraByte = 8,000 boxes = 2,000 file cabinets
- 8 TeraByte = 64,000 boxes = 16,000 file cabinets

www.ArchiveBuilders.com 24003-018

B-218

RAID is a form of ECC

- RAID (**R**edundant **A**rray of **I**nexpensive **D**isks)
- ECC (**E**rror **C**orrecting **C**ode)
- Hot swap disks: can be pulled out while the computer is running
 - Sled (contains the disk)
 - Holster (slot for the sled)
- Accessing data on a disk that has been removed
 - [At the blackboard.]

www.ArchiveBuilders.com 24003-018

B-219

RAID Levels (a Curiosity)

- 0 - striping for speed, no redundancy
- 1 - mirroring for redundancy and duplexing for speed
- 2 - data on one set of disks, ECC on second set of disks
- 3 - separate parity disk, striping for speed
- 4 - like level 2 with a single ECC disk
- 5 - striping of data and striping of parity
- 6 - two levels of parity for more redundancy
- 7 - RAID on a SAN

www.ArchiveBuilders.com 24003-018

B-220

RAID Levels (What is Needed for Document Imaging)

- RAID Level 3 with no striping
 - Each data disk is independent and can be restored independently from CD/DVD if two disks in the RAID array fail simultaneously.
 - This avoids the need to reload the entire RAID array if two disks fail simultaneously.
- Document images have a relatively low transaction rate and low data rate.
 - Disk arrays for document imaging benefit from long RAID strings (many data drives for each parity drive).
 - The long RAID strings increase the probability that two drives will fail simultaneously.

www.ArchiveBuilders.com 24003-018

B-221

Bit Bucket

- Mythical final resting place of used bits.
 - May have originated with the container which accumulated the chaff (punched holes) produced by a keypunch.
- Dirty Bit
 - A flag that indicates that a file has been changed while in RAM and must be rewritten to magnetic disk.
- Sticky Bit
 - A flag that is set with the first use and must be explicitly reset.
- Toggle Bit
 - A flag that switches back and forth, like the caps lock key on a computer keyboard.

www.ArchiveBuilders.com 24003-018

B-222

Optical Disk

www.ArchiveBuilders.com 24003-018

B-223

Optical Disk

- Write Restrictions
- Formats
- HSM (Hierarchical Storage Management)
- Spatial Diversity (7 copies)

www.ArchiveBuilders.com 24003-018

B-224

Optical Disc: Formats

- Spelled 'disc', with a 'c'
 - Following the vinyl 'disc' used in music recording
- CD (**C**ompact **D**isc)
- DVD (**D**igital **V**ideo **D**isc in the vernacular)
 - (Officially just a series of letters)
 - ROM, WORM, RAM
- 5 1/4 Inch (MO, WORM, Phase Change)
- -R (CD-R, DVD-R): a writer for WORM

www.ArchiveBuilders.com 24003-018

B-225

Optical Disc: Write Restrictions

- ROM (**R**ead **O**nly **M**emory)
- WORM (**W**rite **O**nce, **R**ead **M**any)
- RW (**R**ead **W**rite) (rewriteable)
 - Also called RAM (**R**andom **A**ccess **M**emory)
 - MO (**M**agneto **O**ptical) (RAM)
 - Phase Change (RAM)

www.ArchiveBuilders.com 24003-018

B-226

RAM

(**R**andom **A**ccess **M**emory)

www.ArchiveBuilders.com 24003-018

B-227

RAM Recently Dropped to 1/10th its Previous Price, So . . . ?

- Why does everyone instantly need 10 times as much memory in their PCs?
 - Because it costs software vendors money to shrink their software to fit in the available memory. Given the option, they will release the larger, less optimized version of their product.
- With much less expensive memory, it will become common to keep entire databases in RAM.
- [RAM price history on blackboard]

www.ArchiveBuilders.com 24003-018

B-228

Document Management, Document Imaging

Keystrokes vs. RAM Cost

- At 10 keystrokes per second, and \$30.00 per hour of fully burdened data entry operator time, 1 KByte of database entries cost \$1.00.
- At \$1.00 per MegaByte, 1 KByte of database entries cost .1 cent to store.
- Data cost 1 thousand times more to key-in than to store in RAM

www.ArchiveBuilders.com 24003-018

B-229

Preserving the Bits

Security & Integrity

www.ArchiveBuilders.com 24003-018

B-230

Security / Integrity

- Security
 - Protection from Outsiders
- Integrity
 - Protection from System Errors

www.ArchiveBuilders.com 24003-018

B-231

2 Preservation Issues

- Interpreting complex formats
 - If you had the bits, could you interpret them?
 - Covered under imported computer files
- Protecting the bits
 - Covered here
 - Making sure that you do not lose any bits
 - Based on longevity of media and error correcting codes

www.ArchiveBuilders.com 24003-018

B-232

Bit Fade

•Whoa !!!

www.ArchiveBuilders.com 24003-018

B-233

2X, 4X, 6X, 8X, 16X, 32X, 44X

- What is 1X ?
 - A CD can play 74 minutes of music
 - A CD holds 650 MegaBytes of data
 - 650 MegaBytes / 74 minutes =
 - 8.8 MegaBytes / Minute =
 - 70 Megabits / Minute
 - 1.17 Megabits / second \approx 1 Megabit / second
 - 2X = 2 Megabits / second; 4X = 4 Megabits / second
 - 8X = 8 Megabits / second; 16X = 16 Megabits / second
 - At above about 32X the CD drives do not speed up to read the inner track so the constant size pits are read more slowly than 32X on the inner track.

www.ArchiveBuilders.com 24003-018

B-234

Document Management, Document Imaging

Why Some Bits are Badly Written

- You would like your CD / DVD burner to blast one million holes per second in the blank disk — perfectly.
- You want to pay USD \$300 for your burner.
- You want the burner to work flawlessly for 5 years.
- Something has to give.

www.ArchiveBuilders.com 24003-018

B-235

Integrity

- Bit Fading, BER (Bit Error Rate)
 - All bits fade, on communications links, in RAM (Random Access Memory), on magnetic disks, and on optical discs.
 - DLT BER of 10^{**}-17 Corrected, 10^{**}-27 Uncorrected
 - (Digital Linear Tape)
 - CD raw error rate: BER of 10^{**}-4 to 10^{**}-5
 - (1 in 10 thousand to 1 in 100 thousand)
- Parity (Even or Odd)
- CRC (Cyclical Redundancy Check)
- ECC (Error Correcting Codes)
- Digital Signatures, Seals, Encryption

www.ArchiveBuilders.com 24003-018

B-236

Integrity

- Reading the raw error rate
- The ANSI/AIIM MS59-1996 media error monitoring and reporting standard
 - Compliments the ANSI X3.131 media error hardware interface standard
 - Association for *Information and Image Management* / *American National Standards Institute*
- Data can be formatted to avoid losing the redundancy (the ECC) along with the data.
 - Spreading the data out on the media
 - Until it is available, most end users plan to copy CDs and DVDs every 5 years, regardless of the nominal warranty period.

www.ArchiveBuilders.com 24003-018

B-237

Optical Disk: Spatial Diversity

- Volume Copies (The Speed of SneakerNet)
- DVD at Multiple Sites
 - Each Site from a Different Vendor
 - 7 sites recommended
 - (Given the projected cost of USD \$1 DVDs)
- Spatial Diversity Important in Communications
 - Avoid having your backup cable in the same pipe as your primary cable.

www.ArchiveBuilders.com 24003-018

B-238

Will You Lose Your Files?

- Yes - if you have not lost files before
- There is only one way to understand the need for backup
 - Lose some important files
- How serious will the problem be for you?
 - More serious than the last time you lost files
- What can you do?
 - Help your friends
 - If they have a loss, you will share in the moment of truth

www.ArchiveBuilders.com 24003-018

B-239

Tape Backup

- Mammoth 2 from Exabyte.com
- 60 GBytes (Native, Not Compressed)
 - Avoids Purchase of Tape Autochanger for unattended backup
- RAID Tape
 - For fast system backup (over night)
 - Avoids Purchase of Tape Autochanger
- RAID Tape can have a Parity Tape and can therefore be Fault Tolerant

www.ArchiveBuilders.com 24003-018

B-240

Document Management, Document Imaging

Caution: Incompressible Data

- Image data is incompressible because it is already compressed and you cannot compress data twice.
- All tape drive and some communications link capacities are given in compressed form.
- To avoid confusion, always request two numbers, never one number.
 - 1.) The compressed capacity
 - 2.) The uncompressed capacity
 - This parallels the need to always specify 3 numbers in document imaging: the number of pages, the number of documents, and the number of pages per document.

www.ArchiveBuilders.com 24003-018

B-241

Media Precautions

www.ArchiveBuilders.com 24003-018

B-242

Media Precautions

- Magnetic tapes and disks are magnetic.
 - Strong magnetic fields can erase them.
- Magneto optical disks are heated for writing.
 - Strong magnetic fields cannot erase them easily.
- Strong magnetic fields cannot erase optical disks.
- Sunlight can erase exposed optical disks.
- Floppies are destroyed by temperatures over 125 degrees Fahrenheit (50 degrees Celsius - C).
(vs. Fahrenheit 451 for paper. - 233 °C)

www.ArchiveBuilders.com 24003-018

B-243

Short & Long Term Preservation

- Digital copies are perfect because of the ECC.
 - ECC (Error Correcting Code)
- Digital copies can be digital sealed and signed.
 - Preserves Chain of Custody, Integrity of Records
 - Can be checked each morning
- Long term preservation depends exclusively on the longevity of the physical media because a civilization, necessary to copy the records when they begin to fade, may not survive in the long term.

www.ArchiveBuilders.com 24003-018

B-244

Preserving Electronic Records

www.ArchiveBuilders.com 24003-018

B-245

Preserving Electronic Records (1)

- Central Elements in Archiving Data:
 - Electronic Seals
 - Electronic Signatures
 - Encryption
 - Error Correcting Codes
 - Spatial Diversity
 - Emulation of Formats and Hyperlinks (Layer 6)

www.ArchiveBuilders.com 24003-018

B-246

Document Management, Document Imaging

Preserving Electronic Records (2)

- The first tasks of a data archivist would include:
 - Getting the bits and provenance off the incoming media.
 - Affixing the archives' digital seal.
 - Digitally signing the package.
(Archiving is done by people.)

www.ArchiveBuilders.com 24003-018

B-247

How Long Might the Records Last?

- The properly archived digital record will have the potential to last, unchanged, until the end of the universe.
 - Where the end of the universe is defined as that time when the universe collapses (assumed), and we will not be able to tell a one from a zero.
 - There will be no discernable differences

www.ArchiveBuilders.com 24003-018

B-248

Why Digital Copies Are Perfect

- Before a copy is made of a digital record, the ECC (Error Correcting Code) is used to regenerate a perfect copy of the record.
- The copy is then made.
 - The newest copy of a digital record is always the best copy.
 - This is in contradistinction to copies of paper records, where the oldest copy is always the best copy.
- The new copy will last until more bits fade away than the ECC can correct.
 - To preserve the record this must not be allowed to happen.
 - The record must be copied before the bad bits overwhelm the capacity of the ECC to correct any errors.

www.ArchiveBuilders.com 24003-018

B-249

How Often Should You Copy Digital Records?

- As often as your ECC requires.
- ECC's can be made arbitrarily robust.
 - They can protect for as long as desired.
 - But, more robust ECC's are bigger.
 - Use up a larger percentage of the available data storage space for the ECC

www.ArchiveBuilders.com 24003-018

B-250

Administrative Security and Integrity (1)

- Disaster Recovery, Contingency Planning
- Backup, Replication, Mirroring, Failover
- Volume Copies, RAID Disk, RAID Tape (*Redundant Array of Inexpensive Disks*)
- Retention, Archiving, Bit Fading and ECC (*Error Correcting Codes*)

www.ArchiveBuilders.com 24003-018

B-251

Administrative Security and Integrity (2)

- Electronic Signatures, Electronic Seals, Encryption
- Ownership, Copyright, Nondisclosure
- Rules of Evidence, Admissibility in Court

www.ArchiveBuilders.com 24003-018

B-252

Document Management, Document Imaging

Administrative Security and Integrity (3)

- Make 7 copies for spatial diversity
 - To ensure survival of at least one verifiably good copy of every fascicle
 - Also applies to communications and electric power to a lesser extent
- Try to avoid asking people to do something new and different when there is a disaster to recover from.
 - A common error when restoring a backup tape under disaster conditions is to erase the backup tape.

www.ArchiveBuilders.com 24003-018

B-253

Administrative Security and Integrity (4)

- Records management identifies all record series (paper documents) and assigns a retention schedule to them.
 - This can be broadened to identifying all records, hard copy and electronic, in an organization, and managing all record management aspects of those records.

www.ArchiveBuilders.com 24003-018

B-254

Sealing Wax and Other Things

- Sealing wax, from a candle, was dripped on a document, or a ribbon around a document, to seal it until opened by the recipient. An elaborate carved seal was pressed into the hot wax to authenticate the seal.
 - After a wax seal was broken, the document could still be read, but the certainty of the document's authenticity was called into question.
 - This is how electronic signatures work.

www.ArchiveBuilders.com 24003-018

B-255

Changing Data on WORM Discs

- How do you change data on a WORM disc?
 - WORM (Write Once Read Many)
 - Copy the data to a blank disc.
 - Change the data while you are copying it.
- This can be prevented by using digital signatures.

www.ArchiveBuilders.com 24003-018

B-256

Digital Signature: an Electronic Seal

- The Boeing 777 is the world's first paperless aircraft. It was created entirely with digital signatures on electronic documents.
- Digital signatures are required to avoid printing file copies of word processor documents.
- A digital signature is much more than pasting in a raster scanned signature image on a document.

www.ArchiveBuilders.com 24003-018

B-257

Creating a Digital Signature

- To create a digital signature:
 - Create a checksum by adding up the bits in an electronic file.
 - This is like crossfooting the totals on an expense report.
 - Then encrypt the checksum using a personal password.
 - This creates a digitally signed document that cannot be altered without detection.

www.ArchiveBuilders.com 24003-018

B-258

Document Management, Document Imaging

Digital Signature Encryption and Document Access

- Because the document itself need not be encrypted to be digitally signed, it is easy to read the document without the encryption key used for the electronic signature.
 - The digital signature is only used to verify the authenticity of a document.
- Encrypted documents can also be digitally signed.
 - Encrypted documents cannot be read without the encryption key
- Digital signatures last as long as the format, within which they exist, can be interpreted.

www.ArchiveBuilders.com 24003-018

B-259

California and the US Permits Digital Signatures

- The governor of California signed SB 820, (Senate Bill) the Electronic Transactions Act, enabling digital signatures beginning 1/1/2000.
 - <http://www.mercurycenter.com/business/top/044636.htm>
 - California is the first state to recognize the legal validity of electronic signatures.
- The President of the US signed S. 761, (Senate), Electronic Signatures in Global and National Commerce Act on June 30, 2000

www.ArchiveBuilders.com 24003-018

B-260

A System of Digital Signatures

- A system of digital signatures can be used to ensure that the files in a group of files have not been altered and that no files have been added to or removed from the group.
 - Can protect a directory of files
 - Can protect a fascicle
- A system of digital signatures can be used to ensure that no changes have been made to any document in a records center while the documents are under the protection of the records center's chain of custody.
 - Were any document changed over the weekend?
 - Were any directories or fascicles changed over the weekend?

www.ArchiveBuilders.com 24003-018

B-261

Digital Signature Longevity

- Digital signatures last as long as the format, within which they exist, can be interpreted.

www.ArchiveBuilders.com 24003-018

B-262

2 Year Time Horizon: Backup / Disaster Planning

- Because fascicle based system records (on fascicles) do not change, no baseline, and corresponding (complex) incremental backups (and media rotations), are necessary.
 - Will survive a technical staff 2 year time horizon
- All that is required is that as new fascicles are filled, the new fascicles are digitally sealed, and at least 7 duplicate copies of each of the new fascicles are made for offsite storage.
 - The offsite storage should be at 7 or more spatially diverse (different locations) and managerially diverse (different vendors, not under the same management or ownership) sites.
 - Also avoided is a 'keep forever' cycle of baseline backups that is necessary to avoid a worm, where a little of the database is nibbled away, day-by-day (by the software error worm, or by a virus), undetected, over a period of years.

www.ArchiveBuilders.com 24003-018

B-263

A TransFormat (TF) System Fascicle

- A logical box to contain and protect records
- 4.5 GigaBytes of data, to fit on a standard DVD
 - About 80 thousand pages or 8 file cabinets
- Several fascicles will fit on one 73 GigaByte magnetic disk (16 fascicles per 73 GigaByte disk)
- Fascicles are easy for a non-technical system administrator to: watch, find, and restore
- With the right media and ECC (*E*rror *C*orrecting *C*ode), a fascicle can (be engineered to) last forever.

www.ArchiveBuilders.com 24003-018

B-264

Document Management, Document Imaging

Legality

- The US uses English Common Law
 - It is legal unless prohibited
- Many countries use Roman Law
 - It is illegal unless allowed
- In the US, the legality of digital imaging is based on:
 - Is digital imaging the way you conduct your business?
 - Are you using digitizing as a ploy to destroy evidence?
 - For all areas, including all business practices:
 - Legality is established in court, setting precedents.
 - There are many levels of courts, and many courts at each level.
 - A good lawyer can create questions anywhere.

www.ArchiveBuilders.com 24003-018

B-265

Paleoseismology

- The last major earthquake in Southern California was in 1857, The Ft. Tejon Quake
 - Magnitude 8.3
 - The ground slipped 15 to 30 feet (5 to 9 meters)
 - Over a distance of 225 miles (360 Kilometers) from Tulare lake in central California to Cajon Pass.
- The previous 10 major earthquakes occurred in:
AD 671±13, 734±13, 797±22, 997±16, 1048±33, 1100±65, 1346±17, 1480±15, 1812, and 1857
 - Average wait: 132 years (1857 +132 = 1984)
- http://www.cs.csuabak.edu/Geology/Faculty/Negrini/ITF_course/EarthquakeFreq.html
- <http://www.scecdc.sccc.org/forttejo.html>
- <http://nceer.eng.buffalo.edu/faqs/lifeline.html>

www.ArchiveBuilders.com 24003-018

B-266

Hosted Application Web Sites

- ASP (Application Service Provider)
 - 1 second to switch over to backup site
 - With no dynamic data it is simple
 - Operate normally during earthquakes, etc.
 - Can handle 10 times the normal traffic during and after a disaster.
 - Requires a review of the reliability of public Internet access
 - ISP (Internet Service Provider)
 - Satellite access
 - Cell phone access

www.ArchiveBuilders.com 24003-018

B-267

100 TeraByte Data Warehouse

- Using the IBM DB2 database on IBM RS-6000 SP parallel-processing servers
- For Deutsche Telekom, a German telecommunications services company
 - For customer data and call records
 - Used for customer-relationship management
 - 25 TeraBytes completed by 2000Q3
 - 100 TeraBytes system by 2001Q2
- May be the world's largest data warehouse
 - Very few organizations will need 100 TeraBytes (2 billion scanned pages) for document storage

www.ArchiveBuilders.com 24003-018

B-268

Data Warehouse

- Avoiding the inaccessible
 - No more printing for the archive
 - No more microfilm
 - No more tape loading
 - No more jukeboxes and tape silos
- All data online
 - All possible correlations available
 - Access to the entire corporate history

www.ArchiveBuilders.com 24003-018

B-269

Purge

www.ArchiveBuilders.com 24003-018

B-270

Document Management, Document Imaging

Purge

- Magnetic to Optical Migration
- Based on Retention Schedules
- Mixed Purge Dates on Optical Media
- Law Can Be Passed that Say a Deleted Pointer is the Equivalent of Deleted Data
 - Don't Expect That to be Immutable
 - Find the Shred Command

www.ArchiveBuilders.com 24003-018

B-271

DVD

www.ArchiveBuilders.com 24003-018

B-272

DVDs (17 GigaByte CD's)

- DVDs (commonly Digital Video Discs)
- 17 GBytes ROM (Read Only Memory)
 - 2 sided, 2 layers per side
- 7.9 GBytes WORM (Write Once Read Many)
 - And going up to 9.4 GBytes with improvements in technology
- 5.2 GBytes RAM or RW (ReWriteable)

www.ArchiveBuilders.com 24003-018

B-273

Where Will You See It

- 9 Hours of Video (135 min. on each of 4 layers)
- Better Video Than Laser Discs (720 lines)
- 24 Bit Audio (vs 16 Bit on CDs)
- 96 KHz Audio Sampling (vs 44 KHz on CDs)
- 5 Track Surround Sound (vs Obsolete Stereo)
- 8 Language Tracks, 32 Subtitle Tracks
- All Ratings From G to X (via Run Time Editing)

www.ArchiveBuilders.com 24003-018

B-274

Quick Spec

- Same Size as a CD
- Looks Like a CD
- Two Sided
- Two Layers per Side
- Blue Light Special Will Yield 40 GByte ROM

www.ArchiveBuilders.com 24003-018

B-275

CD Only Drives Have Disappeared

- By September 1998, some PC vendors no longer offered CD only CD ROM readers in some product lines.
 - All available PC's in the vendors' product line used a DVD / CD ROM reader.

www.ArchiveBuilders.com 24003-018

B-276

Document Management, Document Imaging

Structural Change

- Soon software will come with a video based tutorial.
 - The DVD will hold the software and the video tutorial.
 - You will be able to load the software from a CD, but you will not get the free tutorial.
 - This will increase training expenses.

www.ArchiveBuilders.com 24003018

B-277

Switching from CD to DVD WORMs

- Switch from CDs to DVDs when DVDs are cheaper
- Blank CD-R (Compact Disc - Recordable)
 - \$1 to \$2
 - Guaranteed for 100 years (plan to copy in 5 years)
 - 1/2 GigaByte (4 records storage boxes) 25 to 50 cents per box
- Blank DVD-R (commonly Digital Video Disc - R)
 - Ditto (in a year or two) (\$1 to \$2, 100 year guarantee)
 - 4.5 GigaBytes (36 records storage boxes) 3 to 5 cents per box

www.ArchiveBuilders.com 24003018

B-278

DVD Establishes New Standards

- In addition to providing a much larger storage format:
 - DVD establishes new standards for:
 - Video
 - Audio
 - Movie formats and subtitles
- DVD Standards not finished yet
 - Many powerful groups are fighting for advantage.

www.ArchiveBuilders.com 24003018

B-279

DVD Pushes PCs

- DVD forces PCs to provide high quality
 - Audio
 - Video
- Making PCs and TVs interchangeable
- The DVD is the last feature needed in PCs that are produced in high volume.
 - When PCs are feature complete, the price will start to drop.

www.ArchiveBuilders.com 24003018

B-280

16X DVD Readers Speed TransFormat System Restore Process

- Coming 1 US dollar per disk blank DVDs
 - 16 X DVD readers (Digital Versatile Disc)
 - cost less than 100 US dollars and can restore a 4.5 Gigabyte fascicle in 10 minutes
- 100 DVD readers
 - Cost of 10 thousand US dollars
 - 450 Gigabytes can be restored in 10 minutes
 - 2.7 Terabytes can be restored in 1 hour

www.ArchiveBuilders.com 24003018

B-281

PC Prices

- Originally based on what people could afford: US\$ 1,500 - US\$ 2,500
- Now that people do not need more:
 - Memory at US\$ 1 per MegaByte
 - 12 GigaByte hard drives
 - More speed than a 10 million transistor 500 MegaHertz Chip
- PC prices will drop to US\$ 10 + a display

www.ArchiveBuilders.com 24003018

B-282

Document Management, Document Imaging

Evolution of Intel Microprocessors 1971 to 2003

| Family | Trade Name (Code Name for Future Chips) | Clock Frequency in MegaHertz** | Millions of Instructions per Second | Date of Introduction | Number of Transistors | Die Size (Pixel Size) | Address Bus Bits |
|--------|---|--------------------------------|-------------------------------------|----------------------|-----------------------|-----------------------|------------------|
| 80786 | (Northwood) | 3,000.0 MHz | TBA | 2003 | TBA | 0.13 micron | 64 bit |
| 80786 | (Madison)**** | TBA | TBA | 2003 | TBA | 0.13 micron | 64 bit |
| 80786 | (Desferald)**** | TBA | TBA | 2002Q2 | TBA | 0.13 micron | 64 bit |
| 80786 | (McKinley)** | 1,000.0 MHz | TBA | 2001Q4 | TBA | 0.18 micron | 64 bit |
| 80786 | (Itanium)** | 800.0 MHz | TBA | 2001Q1 | TBA | 0.18 micron | 64 bit |
| 80686 | Pentium 4***** | 1,500.0 MHz | *1,500.00 MIPS | 2000Q4 | TBA | 0.18 micron | 32 bit |
| 80686 | Pentium III | 1,000.0 MHz | *1,000.00 MIPS | March 1, 2000 | 28.1 million | 0.18 micron | 32 bit |
| 80686 | P III Xeon | 733.0 MHz | *733.00 MIPS | October 25, 1999 | 28.1 million | 0.18 micron | 32 bit |
| 80686 | Mobile P III | 400.0 MHz | *400.00 MIPS | June 14, 1999 | 27.4 million | 0.18 micron | 32 bit |
| 80686 | P III Xeon | 550.0 MHz | *550.00 MIPS | March 17, 1999 | 9.5 million | 0.25 micron | 32 bit |
| 80686 | Pentium III | 500.0 MHz | *500.00 MIPS | February 26, 1999 | 9.5 million | 0.25 micron | 32 bit |
| 80686 | P II Xeon | 400.0 MHz | *400.00 MIPS | June 29, 1998 | 7.5 million | 0.25 micron | 32 bit |
| 80686 | Pentium II | 333.0 MHz | *333.00 MIPS | January 26, 1998 | 7.5 million | 0.25 micron | 32 bit |
| 80686 | Pentium II | 300.0 MHz | *300.00 MIPS | May 7, 1997 | 7.5 million | 0.35 micron | 32 bit |
| 80586 | Pentium Pro | 200.0 MHz | *200.00 MIPS | November 1, 1995 | 5.5 million | 0.35 micron | 32 bit |
| 80586 | Pentium | 133.0 MHz | *133.00 MIPS | June 1995 | 3.3 million | 0.35 micron | 32 bit |
| 80586 | Pentium | 90.0 MHz | *90.00 MIPS | March 7, 1994 | 3.2 million | 0.60 micron | 32 bit |
| 80586 | Pentium | 60.0 MHz | *60.00 MIPS | March 22, 1993 | 3.1 million | 0.80 micron | 32 bit |
| 80486 | 80486 DX2 | 50.0 MHz | *50.00 MIPS | March 3, 1992 | 1.2 million | 0.80 micron | 32 bit |
| 80486 | 486 DX | 25.0 MHz | 20.00 MIPS | April 10, 1989 | 1.2 million | 1.00 micron | 32 bit |
| 80386 | 386 DX | 16.0 MHz | 5.00 MIPS | October 17, 1985 | 275,000 | 1.50 micron | 16 bit |
| 80286 | 80286 | 6.0 MHz | 0.90 MIPS | February 1982 | 134,000 | 1.50 micron | 16 bit |
| 8086 | 8086 | 5.0 MHz | 0.33 MIPS | June 8, 1978 | 29,000 | 3.00 micron | 16 bit |
| 8080 | 8080 | 2.0 MHz | 0.64 MIPS | April 1974 | 6,000 | 6.00 micron | 8 bit |
| 8008 | 8008 | 2 MHz | 0.06 MIPS | April 1972 | 3,500 | 10.00 micron | 8 bit |
| 4004 | 4004 | 1 MHz | 0.06 MIPS | November 15, 1971 | 2,300 | 10.00 micron | 4 bit |

* All note at website below.

www.ArchiveBuilders.com 24003-018

B-283

Chip Real Estate

- Microprocessor vendors are running out of things for the microprocessor chip's 10 million transistors to do.
- The next generation of microprocessor chips will have on the order of 20 million transistors, half of which will be looking for something to do.
- The microprocessor will absorb the function of all the other chips on the mother-board.
- PCs will approach the size of one chip calculators.

www.ArchiveBuilders.com 24003-018

B-284

PCs Become TVs TVs Become PCs

www.ArchiveBuilders.com 24003-018

B-285

All TVs will have a PC on a Chip

- It will be cheaper to put a PC in every TV than to design it out of the TV.
 - It is cheaper to make 100 million of one thing that to make 10 million of 10 different things.
- The PC chip will be merged with the one chip that currently is (operates) a TV.
 - Every PC will include a TV.
 - Every TV will include a PC.

www.ArchiveBuilders.com 24003-018

B-286

Cheap PCs

- \$250 for PC and software within 5 years
 - Microsoft competes with PC hardware vendors for spendable funds.
 - Hardware drops by half every two years
- We have run out of things to want on PCs
 - Microprocessors have 10 million transistors
 - A 128 MegaByte PC has 1 billion transistors
 - A 500 MHz (MegaHertz) PC executes 500 million instructions per second.
 - DVDs in PCs play theater quality movies.

www.ArchiveBuilders.com 24003-018

B-287

Mid 1999 Standard Office PC (1)

- A cusp in the ecology of PC computing
 - aka (also know as): watershed, paradigm shift, big, really big
- New this summer (July, 1999) (all coming at the same time)
 - Under US\$ 2,500.00
 - Windows 98 second edition
 - Office 2000 (Professional and Developer versions)
 - Outlook for your phonebook, calendar, email, and personal contact information.
 - Access database moving to rewritten SQL Server

www.ArchiveBuilders.com 24003-018

B-288

Document Management, Document Imaging

Mid 1999 Standard Office PC (2)

- More:
 - External parallel port CD burners for backup
 - 256 MegaByte RAM (**R**andom **A**ccess **M**emory)
 - 14 GigaByte Magnetic Disk (Hard Drive)
 - 21 inch monitor
 - DVD ROM reader / Surround Sound Movie Player
 - Y2K Millennium bug end run (by total replacement)
 - DSL (**D**igital **S**ubscriber **L**ine) 1/2 to 8 Megabits per second
 - Voice and video phone calls free over the Internet
 - With ATM (**A**synchronous **T**ransfer **M**ode) protocol (next year)
 - Document Scanner

www.ArchiveBuilders.com 24003-018

B-289

Mid 1999 Standard Office PC (3)

- And more:
 - Voice input for word processors, etc.
 - Eliminate RSS (**R**epetitive **S**tress **S**ndrome)
 - Eye tracking to eliminate mouse (next year)
 - Outlook can substitute for Exchange Server if the enterprise has not yet installed Exchange Server
- Coming in 1999 (December 32): Windows 2000
 - Workstation, Small Office, Standard, Enterprise eds.
 - Exchange Server is Outlook for the enterprise.

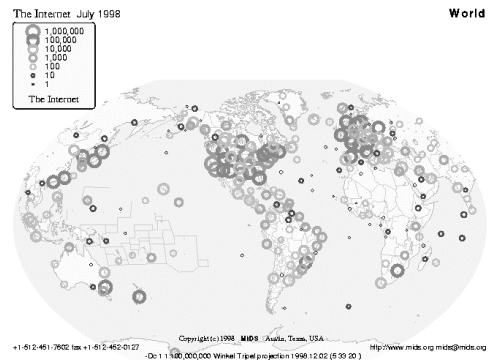
www.ArchiveBuilders.com 24003-018

B-290

Communications

www.ArchiveBuilders.com 24003-018

B-291



40 Million Internet Hosts in Major World Cities
140 million users, 700 million webpages, 206 of 246 countries & territories

www.ArchiveBuilders.com 24003-018

B-292

Communications

- Communications for Document Imaging
- SANs, LANs, WANs, and Telco POTS
- The Internet (and Intranets and Extranets)

www.ArchiveBuilders.com 24003-018

B-293

Internet Delivery – the Big Picture

- January 1999: Internet carries 2 PetaBytes per week
 - 2 thousand TeraBytes, 2 million GigaBytes, 2 billion MegaBytes
 - Doubles every 6 months
 - Free
- DSL can work with existing phone lines at up to 8 Megabits (1 MegaByte) per second. (2 minutes per box)
- ATM : 300 pages per second (5 Boxes per Minute)
- OC192 : 2 file cabinets per second (8 boxes per second)
- Dark Fiber : 4 million boxes per second

www.ArchiveBuilders.com 24003-018

B-294

Document Management, Document Imaging

Offshore Clerks

- Video on demand over the Internet in 3 years.
- Free worldwide video-phone calls.
- Offshore indexing of scanned document
- Offshore receptionists video video-phone
- Offshore system administration
- Offshore Visual Basic programming, etc.

www.ArchiveBuilders.com 24003-018

B-295

Images per Second

www.ArchiveBuilders.com 24003-018

B-296

Internet Delivery

- Modem (56 Kbits/s) : 3 pages per minute
- ISDN (128Kbits/s) : 10 pages per minute (complex to do)
- Cable (TV) Modem (500 Kbits/s): 1 page per second
- DSL (Digital Subscriber Line) (8 Mbits/s): 20 pages/s
- T1 : 3 pages per second
- T3 : 100 pages per second
- LAN (10Base T) 2 pages per second
- LAN (100Base T) 20 pages per second
- ATM : 300 pages per second
- OC192 : 8 boxes per second (2 file cabinets per second)
- Dark Fiber : 4 million boxes per second
– (1 million file cabinets per second)

www.ArchiveBuilders.com 24003-018

B-297

Topologies

www.ArchiveBuilders.com 24003-018

B-298

Topologies

- Ring
- Star
- Bus

www.ArchiveBuilders.com 24003-018

B-299

Geographic Extents

- SANs (Storage / System Area Network)
 - Computer Room
- LANs (Local Area Network)
 - Building
- MANs (Metropolitan Area Network)
 - Campus or Metro Area
 - (Now Becoming Intranets)
- VANs (Value Added Network)
 - World (Becoming Part of the Internet)

www.ArchiveBuilders.com 24003-018

B-300

Document Management, Document Imaging

Internet

www.ArchiveBuilders.com 24003-018

B-301

Images Across the Internet

- [At the Blackboard]
- Internet, Intranet, Extranet, and Trust
- The seven layers of protocols (the protocol stack)
 - Why 7 different protocols are in use simultaneously
 - And, how one protocol can be substituted for another
 - Or, simulated via tunneling
- The structure of the Internet and the protocol stack
 - Why the Internet is called an 'Inter' - 'Net'

www.ArchiveBuilders.com 24003-018

B-302

Protocol Stack

www.ArchiveBuilders.com 24003-018

B-303

Protocol Stack

- Application
- Presentation
- Session
- Transport
- Network
- Link
- Physical

www.ArchiveBuilders.com 24003-018

B-304

Internet Etc.

- **Internet**
 - Made up of linked independent networks
 - Designed to survive nuclear Armageddon
- **Intranet**
 - Private Internet (under owner's control)
 - Can use all internet software and metaphors
- **Extranet**
 - Linked Intranets
 - Linked (shared) security, trust

www.ArchiveBuilders.com 24003-018

B-305

Intranet

- Use the Internet as digital POTS
- Does not allow use by outsiders
- Requires a firewall at every Internet interface
- Will become the universal method of private network data access and transfer

www.ArchiveBuilders.com 24003-018

B-306

Document Management, Document Imaging

Extranet

- Linked Intranets
- Requires cooperative, trusting parties

www.ArchiveBuilders.com 24003-018

B-307

IP (Internet Protocol)

- TCP/IP (Transmission Control Protocol/IP)
- IP Address (Internet Protocol)
 - 192.0.0.0
 - (Base 256)
 - Specifies one of 4 billion possible addresses
 - (2 ** 32)
 - Much too small, being extended to 2 ** 128.

www.ArchiveBuilders.com 24003-018

B-308

ATM

- Asynchronous Transfer Mode
- Required underneath to support smooth (not choppy) personal communications
 - Voice and video telephony
 - Video on demand which will replace:
 - Video stores
 - Cable TV
 - Broadcast TV
 - Security cameras

www.ArchiveBuilders.com 24003-018

B-309

HDTV ATM Studio & WAN Switch

- FORE Systems and Tektronix demonstrated uncompressed HDTV (*H*igh *D*efinition *T*ele*V*ision) switching at NAB 99 (National Association of Broadcasters), in Las Vegas, NV, April 19, 1999 for studio LANs (*L*ocal *A*rea *N*etworks), WANS (*W*ide *A*rea *N*etworks) and the Internet. [www.FORE.com]
- A FORE ATM (*A*synchronous *T*ransfer *M*ode) switch with OC-48c (*O*ptical *C*arrier 2.488 Gigabit/second) ports transported real-time, uncompressed, HDTV in the industry-standard 1.5 Gigabit/second stream. SMPTE292M (*S*ociety of *M*otion *P*icture and *T*ele*V*ision *E*ngineers, SMPTE).
- The Tektronix video edge device is a full 10-bit digital video transport device.
- The FORE ATM Switch is a 40 Gigabit/second, non-blocking, full duplex-switch, capable of switching 20, independent, uncompressed, HDTV signals.
- With this great headroom, (compressed HDTV is much easier to send than uncompressed HDTV), compressed HDTV is very likely to arrive into the home, on demand, over the Internet, on a fiber optic link, within a decade.

www.ArchiveBuilders.com 24003-018

B-310

CSMA/CD (Ethernet)

- CSMA/CD Carrier Sense *M*ultiple *A*ccess with *C*ollision *D*etection

www.ArchiveBuilders.com 24003-018

B-311

Ethernet Segmenting and Routing

- Being replaced by switched ethernet.
- Every NIC (*N*etwork *I*nterface *C*ard) on a LAN can hear everyone else and has to look at their traffic.
- By splitting or segmenting a LAN, you cut the traffic in half on each half of the LAN.
- To get just the messages that need to go across the split, across the split, a router is used.

www.ArchiveBuilders.com 24003-018

B-312

Document Management, Document Imaging

Gateways

- A gateway lets people in one network into the secured environment of another network.
- A firewall is part of an industrial strength gateway.

www.ArchiveBuilders.com 24003-018

B-313

Spatial Diversity

- For communications:
 - Do not route your backup communications link through the same conduit as your primary link.
- For data storage:
 - Do not store your backups next to your computer.

www.ArchiveBuilders.com 24003-018

B-314

Bandwidth

www.ArchiveBuilders.com 24003-018

B-315

Bandwidth

- Bandwidth is literally the frequency range of a transmission medium.
- It is taken to mean the amount of data that can fit through a communications link.
- People can be said to have a fixed bandwidth that can be used to handle minor details (that could be eliminated by consistent system design) or for useful work.

www.ArchiveBuilders.com 24003-018

B-316

A Raster Image is a Digital Analog

- An analog is a replication of something in another medium.
- A raster is a replication of a page in a digital medium.

www.ArchiveBuilders.com 24003-018

B-317

All Digital Signals are Analog

- Digital ones and zeros are mathematical concepts.
- In the real world ones and zeros are voltage or current levels that can be distinguished.
 - We store and transmit our digital analogs using analog digital mediums.
- As electronics technology enters the nano-world of quantum electronics, ones and zeros may be represented by discrete quanta.

www.ArchiveBuilders.com 24003-018

B-318

Document Management, Document Imaging

Finding the Bits

- Recovery of the Clock
 - From an accurate clock (causes slips)
 - Synchronous
 - From a transition in the carrier (Start & Stop Bits)
 - Asynchronous
 - From the signal (limit on the maximum value run)
 - Isochronous

www.ArchiveBuilders.com 24003-018

B-319

Acronyms, Palindromes, and Inverted Eponymy

- Modulator, Demodulator (Modem)
- Radio Detecting and Ranging (Radar)
- Light Amplification by Stimulated Emission of Radiation (Laser)
- Historical Eponymy in Computing
 - Virtual Memory System (for the Virtual Address eXtension computer) for Windows New Technology
 - HAL for International Business Machines

www.ArchiveBuilders.com 24003-018

B-320

C = Frequency x Wave Length

- Microwaves are in the GHz range.
- Fiber Optics Use 1300 nm (*nanometer*) Light
- C is the speed of light in $E = MC^2$
 - C is a universal constant that is a part of the definition of the universe.
- C is about 300 MMeters per second.

www.ArchiveBuilders.com 24003-018

B-321

C = Frequency x Wave Length

- Microwaves:
 - $(10^{11}) \times (3 \times 10^9) = (3 \times 10^{20})$
 - $(100 \text{ mm}) \times (3 \text{ GHz}) = 300,000 \text{ KM} / \text{sec}$
 - C is about 300 MMeters per second = 300,000 KM / sec
- Fiber Optics (1300 Nanometers):
 - $(1.3 \times 10^{-6}) \times (2.3 \times 10^{14}) = (3 \times 10^8)$
 - $(1,300 \text{ nm}) \times (230 \text{ THz}) = 300,000 \text{ KM} / \text{sec}$
- And, a rainbow is one octave

www.ArchiveBuilders.com 24003-018

B-322

Future Networking

www.ArchiveBuilders.com 24003-018

B-323

Meeting the Fiber When It Arrives at Your House

- T1 Was Invented for Rural Areas In the 1960's
- All modem transmissions have been digitized at 64 Kbps for 30 years (even for 300 bit per second modems)
- Cable Modems Are 10 Mbps (*Mega-bits per sec.*)
- DirecTV Channels are 25 Mbps
- OC192 SONET Fiber Interface is 10 Gbps
 - SONET (*S*ynchronous *O*ptical *NET*work)
- Fiber to the Home Will Be Multi-Gigabits per sec.

www.ArchiveBuilders.com 24003-018

B-324

Document Management, Document Imaging

Telco

- Telco (**T**elephone **c**ompany)
- POTS
 - **P**lain **O**ld **T**elephone **S**ervice
- SONET
 - **S**ynchronous **O**ptical **N**ETwork
- Asymmetric Internet
 - Twisted pair to carry mouse clicks (Slow)
 - DirecTV to carry images and video (Fast)

www.ArchiveBuilders.com 24003-018

B-325

Entertaining (Microsoft?)

www.ArchiveBuilders.com 24003-018

B-326

Microsoft Plans for PCs to Host All Forms of Electronic Entertainment

- Interactive, group electronic games
- Education
- Personal communications
- TV and movie viewing
- Radio, audio recordings

www.ArchiveBuilders.com 24003-018

B-327

Dark Fiber

www.ArchiveBuilders.com 24003-018

B-328

Dark Fiber

- Use your own fiber modem.
- Current Max speed is OC192 (10 Gbits/sec.)
- 40 Gbits/sec. with 4X WDM (4 color or wavelength)
(**W**avelength **D**ivision **M**ultiplexing)
- Theoretical Maximum speed is based on the
fiber optic carrier frequency (about 230 THz)
 - NTSC TV is transmitted uncompressed
at about 40 Mbits/sec.
 - 230 Tbits / 40 Mbits =
5 Million NTSC TV channels per fiber

www.ArchiveBuilders.com 24003-018

B-329

Microsoft Internet Strategy

- First: The Internet is Unimportant.
- Second: We will replace the Internet with the
Microsoft Network.
- Currently: We will Embrace the Internet and
all Microsoft Internet Software will be FREE!

www.ArchiveBuilders.com 24003-018

B-330

Document Management, Document Imaging

The Spanish Armada: A Story

- In 1588 the Spanish Armada sailed for England.
- The English spent a considerable sum building a series of signal relay towers from the coast to London.
- The investment was wildly successful. It carried a single bit: "The Spanish are Coming!"
 - If communications costs have been dropping rapidly for hundreds of years, then it should have cost millions to send a single bit in the 16th century.

www.ArchiveBuilders.com 24003-018

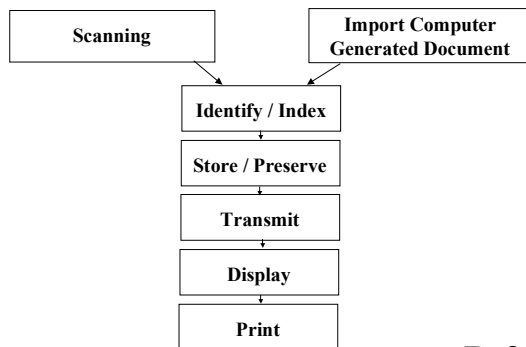
B-331

Thank You

www.ArchiveBuilders.com 24003-018

B-332

Flow



www.ArchiveBuilders.com 24003-018

B-333

Day/Part Slide

Topic

| | | |
|-------|---------|------------------------------------|
| 1 - A | A - 1 | First Principles / Overview |
| 1 - B | A - 59 | Scanning / Printing |
| 1 - C | B - 7 | Importing / Preserving Files |
| 1 - D | B - 186 | Index / Search |
| 2 - A | B - 203 | Store / Purge / Preserve / Display |
| 2 - B | B - 291 | Internet / Systems |
| 2 - C | C - 109 | RFP / Records Management |
| 2 - D | C - 209 | Paradigms, References, & Future |

www.ArchiveBuilders.com 24003-018

B-334

International Association of Aquatic and Marine Science Libraries and Information Centers

IAMSLIC 2000 Conference: Tides of Technology

Workshop: Document Management and Document Imaging for Aquatic and Marine Libraries

September 30 & October 1, 2000, Victoria, British Columbia, Canada

Presented by: Steve Gilheany Tel: +1(310) 937-7000 Fax: +1(310) 937-7001
1209 Manhattan Avenue, C-14, Manhattan Beach, California 90266
<http://www.ArchiveBuilders.com> SteveGilheany@ArchiveBuilders.com

www.ArchiveBuilders.com 24003-018

B-335



IAMSLIC 2000 Conference

Tides of Technology

Workshop:

Document Management and Document Imaging for Aquatic and Marine Libraries

September 30 & October 1, 2000, Victoria, British Columbia, Canada

Presented by: SteveGilheany@ArchiveBuilders.com

www.ArchiveBuilders.com 24003-018

B-336

Document Management, Document Imaging



IAMS LIC 2000 Conference Tides of Technology

Workshop: Document Management and Document Imaging for Aquatic and Marine Libraries

September 30 & October 1, 2000, Victoria, British Columbia, Canada

Presented by: SteveGilheany@ArchiveBuilders.com

www.ArchiveBuilders.com 2240046027

C-1

International Association of Aquatic and Marine
Science Libraries and Information Centers

IAMS LIC 2000 Conference: Tides of Technology

Workshop: Document Management and Document Imaging for Aquatic and Marine Libraries

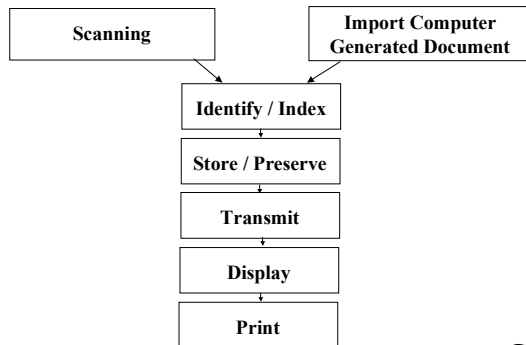
September 30 & October 1, 2000, Victoria, British Columbia, Canada

Presented by: Steve Gilheany Tel: +1(310) 937-7000 Fax: +1(310) 937-7001
1209 Manhattan Avenue, C-14, Manhattan Beach, California 90266
<http://www.ArchiveBuilders.com> SteveGilheany@ArchiveBuilders.com

www.ArchiveBuilders.com 2240046027

C-2

Flow



www.ArchiveBuilders.com 2240046027

C-3

Daily Schedule

Day 1

- 1 - A
- Break
- 1 - B
- Lunch
- 1 - C
- Break
- 1 - D

Day 2

- 2 - A
- Break
- 2 - B
- Lunch
- 2 - C
- Break
- 2 - D

www.ArchiveBuilders.com 2240046027

C-4

| Day/Part | Slide | Topic |
|----------|---------|------------------------------------|
| 1 - A | A - 1 | First Principles / Overview |
| 1 - B | A - 59 | Scanning / Printing |
| 1 - C | B - 7 | Importing / Preserving Files |
| 1 - D | B - 186 | Index / Search |
| 2 - A | B - 203 | Store / Purge / Preserve / Display |
| 2 - B | B - 291 | Internet / Systems |
| 2 - C | C - 109 | RFP / Records Management |
| 2 - D | C - 209 | Paradigms, References, & Future |

www.ArchiveBuilders.com 2240046027

C-5

Workflow

www.ArchiveBuilders.com 2240046027

C-6

Document Management, Document Imaging

Workflow

- Document Routing
- Structured vs Ad Hoc in Same System
- Proprietary vs Commercial Email
- Is Routing Setup Graphical?
- Can You Find In-Process Documents?
 - Even when you delete the node they are enqueued for?

www.ArchiveBuilders.com 224004v027

C-7

Display

www.ArchiveBuilders.com 224004v027

C-8

We Process Vision on Three Channels

- Resolution
 - Text
- Color
 - Is there a red editing mark on this page?
- Motion
 - Instinctive mouse movement to locate cursor
- The channels can be manipulated independently
 - The foundation for optical illusions

www.ArchiveBuilders.com 224004v027

C-9

What Do We See? (Hear?)

- How often is a theater screen completely black?
- Why do people take flash photographs of screens?
- Where does the image on a computer screen come from?
- What does this phrase commonly mean: “To see the the light?”.
- Why does reading backwards help?
 - Answer: Start here, read back four lines.
- Why is a blind spot not a spot?
 - Because it does not exist.
 - It is outside a person’s universe, perception, paradigm.
 - As a spot it fits in a one paradigm paradigm.

www.ArchiveBuilders.com 224004v027

C-10

Displays

- 21 inch CRTs (Cathode Ray Tubes) are under USD \$1 Thousand.
 - 80 Hz refresh rate.
 - Heavy, bulky, hard to use on a public counter
 - Users with bi-focals can only see a resolution of 1280 x 1024
- 18 inch flat panel displays
 - Under USD \$2 thousand and dropping.
- 20 inch flat panel displays are coming down in price.
 - Larger sizes are coming
- 16 to 9 flat panel displays are available.
 - (HDTV aspect ratio)
- Black and White displays have generally disappeared.

www.ArchiveBuilders.com 224004v027

C-11

Common Display Resolutions

| | | |
|-------------|-------|-----------------------------|
| 320 x 200 | CGA | Color Graphics Adapter |
| 640 x 350 | EGA | Enhanced Graphics Adapter |
| 640 x 480 | VGA | Video Graphics Adapter |
| 800 x 600 | SVGA | Super VGA |
| 1024 x 768 | XGA | eXtended Graphics Adapter |
| 1152 x 870 | Mac | 72 dpi |
| 1152 x 900 | Sun | 72 dpi |
| 1280 x 1024 | UVGA | Ultra VGA |
| 1600 x 1200 | UXGA | Ultra XGA |
| 1920 x 1200 | HDTV | Computer High Definition TV |
| 1800 x 1400 | Color | High Resolution Document |
| 2048 x 1536 | B & W | High Resolution Document |

www.ArchiveBuilders.com 224004v027

C-12

Document Management, Document Imaging

Pixels per Inch vs Pixels per Image

www.ArchiveBuilders.com 224004v027

C-13

Pixels per Inch / Pixels per Image

- Pixels per Inch
 - Used in scanned and printed documents
 - Number of pixels depends on the image size
- Pixels per Image
 - Used in displays and scanned photographs
 - Number of pixels is fixed
 - Number of pixels is independent of image size
 - Used in television (smaller screen appears sharper)

www.ArchiveBuilders.com 224004v027

C-14

Kodak PhotoCD Image Sizes

- 1/16 base = 128 x 192 [2 x 3][2**6 x 2**6]
 - (thumbnail, index print on CD cover)
- 1/4 base = 256 x 384 [2 x 3][2**7 x 2**7]
 - (largest Kodak size that is smaller than 480 x 640 for display on TV)
- 1 base = 512 x 768 [2 x 3][2**8 x 2**8]
- 4 base = 1024 x 1536 [2 x 3][2**9 x 2**9]
 - (largest Kodak size that is smaller than 1920 x 1152 for HDTV)
- 16 base = 2048 x 3072 [2 x 3][2**10 x 2**10]
 - (captures all the resolution on most 35 mm film images)
- 64 base = 4096 x 6144 [2 x 3][2**11 x 2**11]
 - (for most film formats larger than 35 mm)

www.ArchiveBuilders.com 224004v027

C-15

Number of Pixels Displayed

- Computer Display
 - 21 inch (525 mm), 1600 by 1200 pixel resolution
 - Displaying an 8 1/2 by 11 inch letter size page
- For a 300 dpi page:
 - How many of the scanned pixels are displayed?
 - What percentage?
 - How many pixels seem to be displayed?
 - Scale to gray

www.ArchiveBuilders.com 224004v027

C-16

GIF and Color Compression

- Based on a color look-up table
 - Invented to reduce V-RAM requirements
 - (Video Random Access Memory)
 - Table has 256 (8 address bits) 24 bit (full color) entries
 - 8 bits are required for each pixel on the display
 - Table runs at the pixel rate of the display
 - (up to 300 MegaHertz)

www.ArchiveBuilders.com 224004v027

C-17

Accommodating Disabilities

- Librarians, archivists, and museum curators have a professional responsibility to accommodate disabilities.
- Systems should be created that produce individualized color shifts for each colorblind patron so that each person can see images with the best possible color fidelity.
 - 256, thousands, and millions of colors display controllers are examples of such a color mapping, as are GIF images.
 - User profiles can have a color perception map added by simply presenting a color perception test to users at their first logon.
 - When necessary, the provenance of a displayed image can include a color space transformation to match the color gambit available to each color blind patron.

www.ArchiveBuilders.com 224004v027

C-18

Document Management, Document Imaging

Windows 98 and Displays

- Microsoft Windows 98 Supports up to 8 Displays
 - Useful for Cubic Surround Vision (Especially Games)
 - Front (3), Back, Left, Right, Above, Below
- For Document Imaging
 - A 1280 x 1024 Color Monitor for Command Windows
 - A 2048 x 1500 Black & White Monitor for Heads Down Work
 - (But, black and white displays have virtually disappeared.)

www.ArchiveBuilders.com 224004v027

C-19

Viewing Features / Capabilities

- Multiple Documents
- Multiple Pages
- Multiple Layers (Drawings, GIS)
- Reference Window
- Pan, Zoom, Scroll, Roam
- Table of Contents
- Measurement (Drawings, GIS)
- Annotation
- Redline
- View Length and Width of Page
- Thumbnails
- Magnifier

www.ArchiveBuilders.com 224004v027

C-20

Edit

www.ArchiveBuilders.com 224004v027

C-21

Document Imaging System Features

- Import/Export Formats Supported
- Annotation (no standard for export)
[Most from Spicer and Watermark]

www.ArchiveBuilders.com 224004v027

C-22

Image Clean Up

- Bates Numbers
- Rotate
- Deskew
- Invert Black/White
- Despeckle
- Mirror
- Crop
- Merge
- Resize
- OCR

www.ArchiveBuilders.com 224004v027

C-23

Edit Feature Settable Attributes

- Color
- Font
- Fill Types
- Line Styles
- Line Thickness
- Object Placement
- Polygon Cut/Copy/Paste

www.ArchiveBuilders.com 224004v027

C-24

Document Management, Document Imaging

Markup / Annotate

- Sticky Note
- Circle
- Line
- Rotate
- ActiveX Object
- Copy
- Modify
- Rubout
- Sound
- Cut
- Move
- Shader
- Drag and Drop
- Delete
- Paste
- Sketch
- Annotate Text
- Dimension
- Polygon
- Symbols
- Arrow
- Ellipse
- Polyline
- Text
- Erase
- Redact Black/White
- Unbind
- bind
- Highlighter
- Redo
- Undo
- Box
- Hotspots
- Resize

www.ArchiveBuilders.com 224004v027

C-25

Brought to you by the number



www.ArchiveBuilders.com 224004v027

C-26

The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information

- My problem is that I have been persecuted by an integer. For seven years this number has followed me around . . . ” George A. Miller (1956)
 - <http://cogprints.soton.ac.uk/documents/disk0/00/00/07/30/cog00000730-00/miller.html>
 - *The Psychological Review*, 1956, 63:81-97

www.ArchiveBuilders.com 224004v027

C-27

The Magical Number 7

- Beyond a certain point, people and organizations give up and abandon systems
 - a person can consider a maximum of seven different things at one time.
 - If an eight item is added, failure occurs.
- By grouping items, more complexity can be handled.
 - By grouping 49 elements into 7 sets of 7 elements each, the larger problem can be made tractable.
 - We can consider 7 groups at one time, but we cannot consider 49 elements, all at the same time.

www.ArchiveBuilders.com 224004v027

C-28

The Magical Number 7 (cont.)

- The problem is that there is a large stair step of effort required (to group elements) every time the number of elements, or groups of elements, exceeds seven.
 - It is this stair step that is avoided by abandoning a TF system (or any system).
 - Before the systems complexity requires making the effort to ascend to the next level of grouping.

www.ArchiveBuilders.com 224004v027

C-29

The Magical Number 7 (cont.)

- This limit of seven is also the basis for span of control.
 - The number people a person can manage
 - And: the height of an organization’s pyramid is determined by:
 - The number of employees
 - The complexity of the work being done at each level.

•Is Pervasive

www.ArchiveBuilders.com 224004v027

C-30

Document Management, Document Imaging

System Design

www.ArchiveBuilders.com 224004v027

C-31

Change

- Are you subconsciously planning for the Internet to stay pretty much the way it is today?
 - Today a single fiber can carry 320 Gbits (40 GBytes) per second
- Are you planning your systems for what was **or for what will be?**
- A 256 MByte computer has 2 billion transistors
 - Is this twice as good as a 1 billion transistor computer?
 - Weren't 9 transistor radios better than 7 transistor radios?

www.ArchiveBuilders.com 224004v027

C-32

Costs that Disappear from Our Radar Screen

- Soon, some costs will not be a consideration
 - The cost of storage
 - The cost of transmission
- Has this happened before?
 - The cost of a glass of water
 - The cost of Internet messages
 - [Digital watches]
 - A US \$1 railroad ticket from St. Louis to LA
 - How do we know costs will not go back up?

www.ArchiveBuilders.com 224004v027

C-33

Preliminary System Sizing Exercise

www.ArchiveBuilders.com 224004v027

C-34

System Design Sizing Assumptions

- 1 Page = 50 KBytes when Stored
- 1 File Cabinet (4 Drawer) = 10,000 Pages
- 1 File Cabinet = 1 CD ROM (1/2 GByte)

www.ArchiveBuilders.com 224004v027

C-35

Storage Required (and Cost)

www.ArchiveBuilders.com 224004v027

C-36

Document Management, Document Imaging

System Design

- Kick the Tires
- Get Demos
- Listen to Vendors
- Read Literature
- Visit Working Sites, Ask Questions
- Try a Pilot
- Buy a Scanner for Your Home PC

www.ArchiveBuilders.com 224004v027

C-37

Incorporating Existing Equipment

- Can easily cost more than buying the new equipment.
- Find a way to move the old equipment down into the organization.

www.ArchiveBuilders.com 224004v027

C-38

Asking Good Questions

- If you ask good questions, vendors will assign better people to your project.
 - Vendors will know they can't get away with less.
 - Successful projects are better than unsuccessful ones.
 - But vendors must plan for many unsuccessful projects.
 - Vendors are hungry for great reference sites.
 - Success requires support from great customers.

www.ArchiveBuilders.com 224004v027

C-39

Technical Relationships

- Calculate the crossover between optical and magnetic storage for your system. (yourself)
- Consider service bureaus for scanning your backfile (especially odd sizes like drawings).
- Distinguish between scanned paper images and imported computer files. (WP, CAD, etc.)
- Learn about electronic signatures, electronic seals, and encryption (my favorites).

www.ArchiveBuilders.com 224004v027

C-40

Possible Integration Requests

- All hardware software, hardware, integration, and maintenance services must be bid as a single turn-key installation package.
- The system integrator must act as prime contractor and take full responsibility for all system components: hardware, software, and integration, during the system design, installation, and maintenance period.

www.ArchiveBuilders.com 224004v027

C-41

What You Say is Not What They Hear

- A person, from outside the United States, was on vacation in the US. He was traveling on a very poor road. He saw a sign saying "End of Pavement". Assuming he was about to come upon the start of good pavement, he sped up.
- He should have known we speak differently, because in his brand of English it is illegal to drive on the pavements.

www.ArchiveBuilders.com 224004v027

C-42

Document Management, Document Imaging

Assembly into a System

www.ArchiveBuilders.com 224004v027

C-43

Network

- Windows 2000 has free networking.
- 100BaseT is coming down in price.
- Neatness counts in network and security configuration.
- Watch for bottlenecks.

www.ArchiveBuilders.com 224004v027

C-44

Server (1)

- Compaq works better with Compaq components.
- Choose a PC designed to be a server
- Get a PC that is as fast as possible.
- Manage price as a percent of system cost.
- Calculate the crossover price point from optical to magnetic storage.

www.ArchiveBuilders.com 224004v027

C-45

Server (2)

- A 5 disk hardware RAID array is easier to understand than 30 disk software RAID.
- Plan to make 7 copies of all files on CD's or DVD's and send at least two copies out of country (state).
- CD's require a CD-R (A CD Recorder). There are also DVD-R's.

www.ArchiveBuilders.com 224004v027

C-46

Client

- One step down from 'as fast as possible' (decompression is in software)
- If necessary people can live with 17 or even 15 inch monitors.
- And with 80 or even 75 Hz refresh rates. (but not lower)
- Black and white can work, but color is more compatible with most GUI software features.
- Spend \$200, get a 10 Gig disk.

www.ArchiveBuilders.com 224004v027

C-47

System Design Considerations

www.ArchiveBuilders.com 224004v027

C-48

Document Management, Document Imaging

Scaleability

- The scaleability of most document imaging systems is tied to the scaleability of Windows 2000.
- As Windows 2000 scales up and out, so do the document imaging systems.

www.ArchiveBuilders.com 224004v027

C-49

Standards

- Allow document interchange
- Allow migration from your vendor
- Industry standards (defacto)
- Official standards (voluntary in USA)

www.ArchiveBuilders.com 224004v027

C-50

Bottlenecks

- There is always a bottleneck just below the surface of calm system water.
- Removing a bottleneck always reveals the next bottleneck.

www.ArchiveBuilders.com 224004v027

C-51

Safety Factors

- The smaller the system the larger the safety factor.
- Below the RFP limit, at least 100 percent, 250 percent is fine.
- A pilot will allow cutting the safety factor in half.
- Three years of operation will let the sharp pencil people in.

www.ArchiveBuilders.com 224004v027

C-52

Stockpiling Equipment

- **Don't.**
- There is no component of document imaging equipment that does not have a high evaporation rate.

www.ArchiveBuilders.com 224004v027

C-53

Uptime (/ Downtime)

- 98 Percent is reasonable for computer systems
 - (98 percent uptime is 3 hours downtime per week.)
 - If you included the words 'planned downtime.'
- Telephone COs (Central Offices) are planned to have one hour of downtime in 40 years.
 - This includes operator error.
 - On average, COs in the US are doing better than this.
- Frequently included under the general term 'good practices.'

www.ArchiveBuilders.com 224004v027

C-54

Document Management, Document Imaging

Flip Rate

- We can see 60 pages per second
- Imagine looking for a word in a dictionary
 - You flip through 100 pages in 1 to 2 seconds
 - The art of book design makes this possible
 - Book design evolved to exploit our visual system
- Computer system designers think 1 page per second is good, 5 pages per second is overkill
 - This is why computer systems are not as fast as books.
 - And why computer system designers have so much paper at their desks.

www.ArchiveBuilders.com 224004v027

C-55

When You Pick Up a Phone

- You expect dialtone.
 - You don't want a dialtone, you expect a dialtone.
 - You plan your life based on having a dialtone.
- Dialtone is slow after an earthquake.
 - We think the phone is broken.
- In some countries waiting for a dialtone is a full time job.

www.ArchiveBuilders.com 224004v027

C-56

The Nines of Reliability

- One nine = 1 hour per day
- Two nines = 1 hour per week
- Three nines = 1 hour per six weeks
- Four nines = 1 hour per year
- Five nines = 1 hour per decade
- Six nines = 1 hour per century
- The CO (**C**entral **O**ffice) goal is about 6 nines
 - The 6 thousand COs in the US are doing better than that.

www.ArchiveBuilders.com 224004v027

C-57

The 1994 San Francisco Quake

- The telephones worked
 - The whole time
 - No reports of problems
 - All the COs ran on emergency generators for weeks.
- What happened after 24 hours?
 - The fuel tanks were empty, there was no power to pump diesel fuel, the diesel fuel distributors in the Bay Area were not interested in diesel fuel, just their families.
 - The phones worked.

www.ArchiveBuilders.com 224004v027

C-58

Image Enabled Applications

- Linking document images with your existing databases
- Requires a degree of programming sophistication
- Will become the standard for all applications in the future
 - Until paper disappears

www.ArchiveBuilders.com 224004v027

C-59

Retrieval / Indexing

- How Do You Do It Now?
 - Minimizes Training
 - Lowers Cost to Implement
- What Is the Most Needed New Capability?
 - What Would Pay for the System in a Month?
 - What Would Be Almost Free,
But Have High Value?

www.ArchiveBuilders.com 224004v027

C-60

Document Management, Document Imaging

Document Indexing

- In the Existing Database
- Input Indexing
- OCR
- Other

www.ArchiveBuilders.com 224004v027

C-61

Making It Work

- William Mulholland ran out of money while building the Los Angeles Aqueduct.
- He only got to the north end of the San Fernando Valley.
- The bonds would go into default if the aqueduct did not reach the City of Los Angeles.
- By having the City of Los Angeles annex the San Fernando Valley, the aqueduct reached the City of Los Angeles with no further work.

www.ArchiveBuilders.com 224004v027

C-62

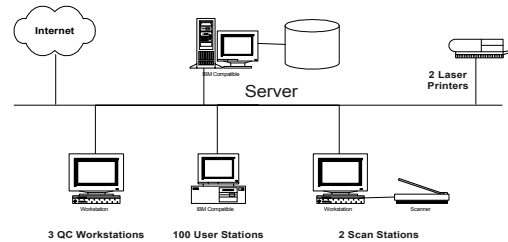
Draw a Nice Picture

- Pizzazz sells.

www.ArchiveBuilders.com 224004v027

C-63

Network Diagram



www.ArchiveBuilders.com 224004v027

C-64

Managing a Technical Project (For a Non-Technical Manager)

www.ArchiveBuilders.com 224004v027

C-65

The Decisions are Yours

The Numbers
Won't Select Your System
for You

www.ArchiveBuilders.com 224004v027

C-66

Document Management, Document Imaging

But . . .

- You Must Do Your Homework
 - Try Everything Yourself as Training
 - Offer to Help Others With Their Projects
 - Then Ask the Experts
 - Background Helps When Managing a Project

www.ArchiveBuilders.com 224004v027

C-67

Your System Selection Will Be Based on Trust

- You Will Trust That:
 - You Selected the Right Criteria
 - You Estimated Correctly
 - Your Vendor Can Deliver
 - Your Vendor Will Deliver
 - Your Organization Will Support You Throughout Your Project

www.ArchiveBuilders.com 224004v027

C-68

Question Everything

- Question People Who Agree with You
- Question People Who Disagree with You

www.ArchiveBuilders.com 224004v027

C-69

Question Everything

Then Make Decisions

www.ArchiveBuilders.com 224004v027

C-70

Invest in Trust

- Find People You Like
- Build a Network of Support
 - And . . .
- Do Your Homework

www.ArchiveBuilders.com 224004v027

C-71

What to Do?

- Make Friends
- Form a Network of Support
- Do Your Homework

www.ArchiveBuilders.com 224004v027

C-72

Document Management, Document Imaging

More Details

- Microsoft's document management product
 - Tahoe with Windows 2002
- GASB 34 (Governmental Accounting Standards Board)
 - Extension to cost of lost records
- Single fibers that can transmit a Petabit per second
 - Provide, without repeaters, on each fiber, 1 billion T-1-like video channels of 1 megabit per second each
 - in development: <http://www.Omni-Guide.com>
- Disaster plan
 - No change in system operation during and after disasters.
 - Multiple servers at ASPs on the Internet (Application Service Providers)
- World defining paradigms of paradigms

www.ArchiveBuilders.com 224004v027

C-73

More Details: Microsoft

- **Microsoft** document imaging system (Tahoe) 2001
 - Industry consolidation
 - Larger market for TF records management systems
- Window ME manages document and image scanners
 - ME (*Millennium Edition*) folded into Windows 2002
- Some Microsoft products are free with many computer purchases
 - Word and Outlook (and perhaps portions of Tahoe)
- Microsoft Windows ME and Windows 2000 will merge in Windows 2002 (projected designation) in the fall of 2001
 - Having one OS (Operating System) to develop to will help all developers, most especially Microsoft itself.
- Microsoft has a new draw, paint, photo editing program called PhotoDraw.

www.ArchiveBuilders.com 224004v027

C-74

GASB 34

- *Basic Financial Statements—and Management's Discussion and Analysis—for (US) State and Local Governments (Statement Number 34) (June 1999)*
 - <http://www.Rutgers.edu/Accounting/raw/gasb/repmodel/index.html>
 - Require government entities to account for value of physical infrastructure, including state of repair, and quality of maintenance.
- Could be extended to include quality of records and quality of records management
 - Lost records can lower the value of 1 billion US dollars of infrastructure by 10 percent (reduction may vary)
 - Due to uncertain location, uncertain construction standards
 - A reduction of 100 million US dollars because of the decreased value of infrastructure caused by poor records or lost records.

www.ArchiveBuilders.com 224004v027

C-75

The Environment

www.ArchiveBuilders.com 224004v027

C-76

Microsoft

(And the Seven Dwarfs, formerly the IBM BUNCH)
(BUNCH: Burroughs, Univac, NCR, Control Data, Honeywell) (& RCA, GE)
(And Steve Ballmer, President)

www.ArchiveBuilders.com 224004v027

C-77

Environment: Why Microsoft?

- Microsoft, to some extent, defines the computer industry.
- Microsoft defines what is available technically to support system implementation.
- Microsoft watching provides input for future business plans and models.

www.ArchiveBuilders.com 224004v027

C-78

Document Management, Document Imaging

How to Evaluate an OS (Operating System)

- UNIX people evaluate OS's on technical merit.
- Accountants and Managers buy computers.
 - Does the software create a fungible staff?
 - Will the software last a long time?
- Only if desperate for an absolutely required (technically required) feature, will the UNIX person enter into the procurement process.

www.ArchiveBuilders.com 2240046027

C-79

Me . . . ? Evaluate an OS?

- Wouldn't you just buy Windows?
 - Yes, most people would, and do.
- Microsoft is a Master of Paradigm Marketing
 - First Microsoft sold everyone on the idea (Paradigm) that Microsoft is the place to buy software.
 - Now, when most people look for an OS or Office Application, they only look at what Microsoft has.

www.ArchiveBuilders.com 2240046027

C-80

What is Microsoft Up To?

www.ArchiveBuilders.com 2240046027

C-81

8 Billion Dollars (USD) in Profits per Year

(And USD \$3 Billion in R&D)

www.ArchiveBuilders.com 2240046027

C-82

Microsoft

- US\$ 2 Billion per quarter profit
- Owns the foundation of the information economy
 - Won years ago
 - Is expanding in every direction
- Benefits the most from increases in the stability and structure of its own products
- Has what customers want most
 - Stability
 - Predictability
 - Ubiquity

www.ArchiveBuilders.com 2240046027

C-83

USD \$500 Billion Capitalization

(Microsoft is now the
world's largest company)

www.ArchiveBuilders.com 2240046027

C-84

Document Management, Document Imaging

How Can I Prove Windows [2000] Will Last 30 Years?

Because in 30 years,
Bill Gates will be 75!

www.ArchiveBuilders.com 224004v027

C-85

Limits on Microsoft

- Microsoft owns the foundation of the information economy.
 - Microsoft can move into a new area of its domain at any time.
 - Microsoft's two year old Expedia travel service does 16 million dollars (US) of business per week.
 - The information economy is said to be 70 to 90 percent of the world economy.
 - The world economy is about \$20 Trillion US.
 - Microsoft's share is about \$15 Trillion US.

www.ArchiveBuilders.com 224004v027

C-86

Unicode

The foundation for unified, single worldwide software releases.

www.ArchiveBuilders.com 224004v027

C-87

Difference Between Microsoft and Linux

- Microsoft increases responsiveness based on purchasing power
- Linux increases responsiveness based on technical expertise and devotion to the concept of Linux

www.ArchiveBuilders.com 224004v027

C-88

What Does Microsoft Have?

- Ubiquity
- The same thing that a marching army has:
 - Many small steps taken in unison are unstoppable.
 - A marching army must break step to avoid tearing apart suspension bridges.
- Microsoft unifies the industry.
 - Microsoft provides a stable development platform.
 - Microsoft delivers a huge, homogeneous market.
- Microsoft builds acceptable products, as needed.

www.ArchiveBuilders.com 224004v027

C-89

Merging Applications and the Operating System

www.ArchiveBuilders.com 224004v027

C-90

Document Management, Document Imaging

Merging Windows 2001, SQL Server, and Exchange

- Microsoft plans to be the first company to take advantage of the freedoms afforded by its object oriented architecture.
- Steve Ballmer, President of Microsoft, speaking in Long Beach, California, Microsoft TechNet Briefing, (Microsoft event number 25316, <http://events.Microsoft.com>), on August 12, 1999, said:
- In the release (both Windows and applications) after the 2000 release this year, Microsoft planned to merge the Windows 2000 file system, the Microsoft SQL Server database, and Exchange (the equivalent of Outlook on servers).

www.ArchiveBuilders.com 224004v027

C-91

Groupware

- Software to support collaboration
- Now means Microsoft Exchange
 - Currently is a mail program
 - Microsoft announced groupware components in Sept. 1998
 - All files are full text indexed as part of the file system
 - Will expand to all documents
 - Microsoft Word stores raster images now
 - The active directory file system of Windows 2000 in 1999 will provide document structure as part of the file system
 - Microsoft recommends that organizations maintain their organization charts and telephone books in Exchange

www.ArchiveBuilders.com 224004v027

C-92

Clusters

www.ArchiveBuilders.com 224004v027

C-93

Clusters of Commodity Components are Economical

- Used for scaling out (a substitute for scaling up)
- Clusters can be built with high-volume components which are relatively inexpensive.
- Inexpensive clusters can be built with commodity four-way SMP hardware systems.
- Because they are built from commodity components, commodity clusters can grow in smaller increments.
- You can add disks or nodes or network cards as needed rather than having to buy a huge new box each time you grow.

www.ArchiveBuilders.com 224004v027

C-94

Reliability

www.ArchiveBuilders.com 224004v027

C-95

Failover Clustering in SQL Server

- Microsoft SQL Server 7.0, in cooperation with various hardware vendors, supports high-availability databases via cluster based fail-over from one node to another.
- Active-X Transactions allow Microsoft SQL Server databases to be partitioned among multiple Windows 2000 Servers.
- The Microsoft SQL Enterprise Manager allows an operator to monitor and manage multiple SQL Servers from a single console.

www.ArchiveBuilders.com 224004v027

C-96

Document Management, Document Imaging

Continuous Availability

- The cost of downtime increases with system size.
- Paradoxically, larger systems have more components and so are more likely to break.
- Scalable systems must deliver continuous 7 x 24 availability (7 days x 24 hours per day).
 - They must allow online operations (backup, recovery, reorganization, upgrades).
 - Must mask hardware and software failures with automatic fail-over to another server.

www.ArchiveBuilders.com 224004v027

C-97

Novell (1)

- Started to help Microsoft sell into networking opportunities
 - When Novell makes a sale, it increases Microsoft's overall market size and sells more Windows products
- Created a large market base for Microsoft networking products
 - When Microsoft sells networking it increases Microsoft's revenue
- Microsoft Networking is free with W2K

www.ArchiveBuilders.com 224004v027

C-98

Novell (2)

- It is currently more profitable for Windows 2000 to be compatible with Novell
 - Compatibility was announced in making it so
 - and then followed by implementation efforts by users and other vendors
- Eventually it may be more profitable for W2K to not be compatible with Novell
 - Incompatibility can be announced to make it so
- The newest version of Lotus Notes does not support Novell

www.ArchiveBuilders.com 224004v027

C-99

Friendly

www.ArchiveBuilders.com 224004v027

C-100

PC Advances to Provide an "Appliance" Experience

- Extremely simple operation (Appliance-ness)
- Anyone can do the top 10 tasks immediately with no prior experience with device; tasks include:
 - Playing a game
 - Watching a movie
 - Writing an e-mail, letter, or invitation
 - Browsing the Internet
 - Hooking up another device
 - Listening to voice messages

www.ArchiveBuilders.com 224004v027

C-101

Responsive

www.ArchiveBuilders.com 224004v027

C-102

Document Management, Document Imaging

Microsoft: “For the PC to become a more integral part of daily life in the office and the home, the PC must be instantly available to answer the phone, display new e-mail, or browse the Internet. As with appliances, the PC must always be on and ready for use but appear to be off when not in use. The PC hardware and software must be capable of responding immediately to the On button, network or communication events, and other actions. Finally, the PC must be capable of returning to its “off but ready” state automatically.”

www.ArchiveBuilders.com 224004v027

C-103

Forgiving

www.ArchiveBuilders.com 224004v027

C-104

Microsoft: “Device bay devices must be able to survive harsh environments. Examples of harsh environments include being carried inside a soft-sided briefcase, device insertion by a toddler, or user attempts to insert devices in the wrong orientation with considerable force.”

www.ArchiveBuilders.com 224004v027

C-105

Entertaining

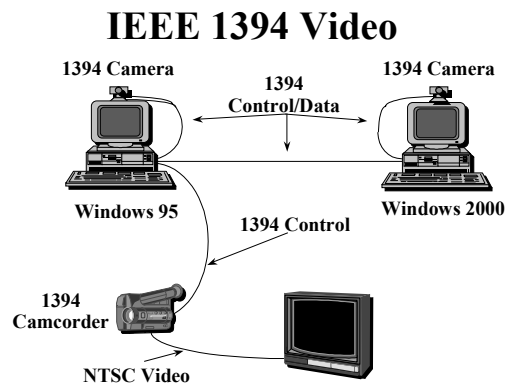
www.ArchiveBuilders.com 224004v027

C-106

Microsoft: “Entertainment PC 97 is an interactive multimedia system optimized for games, education, personal communications, and TV viewing. Key features that differentiate the Entertainment PC 97 from other system types include 3-D graphics and audio, MPEG-II support, high-performance video [DVD], high-fidelity audio, and connections to consumer electronics devices such as home stereos, telephones, VCRs, camcorders, digital satellite systems, and so on.”

www.ArchiveBuilders.com 224004v027

C-107

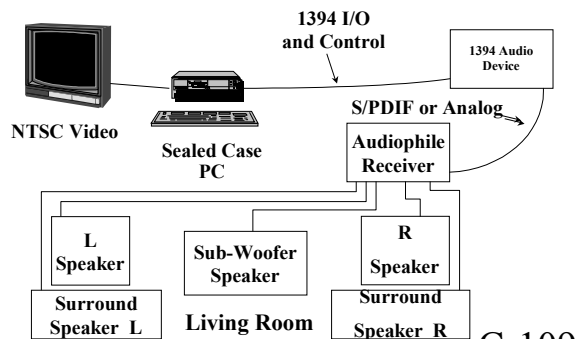


www.ArchiveBuilders.com 224004v027

C-108

Document Management, Document Imaging

PC in the Living Room



www.ArchiveBuilders.com 2240046027

C-109

Records Management

www.ArchiveBuilders.com 2240046027

C-110

Records Management

- Views the Recorded Activity of the Corporation as a Whole
- Assesses All Constituencies' Needs and Requirement for the Records
- Reviews Records for Security, Integrity, and Accessibility

www.ArchiveBuilders.com 2240046027

C-111

Records Managers in the Third Millennium (The View from Inside)

- Professionalization
 - What was accidental is now a profession
- Diversification
 - Records now include voice and email, databases, and more
 - Knowledge Management
- Integration
 - Records managers work with IS and top management
- Background of technology
 - Technological change continues to accelerate
- Education is the key
 - Professional organizations: ARMA
 - Management, marketing, software

www.ArchiveBuilders.com 2240046027

C-112

Records Management Time Frame

- Records Managers Must Accommodate Changes in Technology from the Time a Record is Created Until the Record is Destroyed

www.ArchiveBuilders.com 2240046027

C-113

CD ROM's Are 15 Years Old

www.ArchiveBuilders.com 2240046027

C-114

Document Management, Document Imaging

The Internet is 30 Years Old

www.ArchiveBuilders.com 224004v027

C-115

The Third Millennium Started in

1994

www.ArchiveBuilders.com 224004v027

C-116

Stephen the Short Was Off by 6 Years When He Established the Calendar for the Catholic Church.

www.ArchiveBuilders.com 224004v027

C-117

**We Base A Lot
on the Records We Keep**

www.ArchiveBuilders.com 224004v027

C-118

History of Records Management (1)

- Dr. Nathaniel S. Rousenau invented vertical filing (File Cabinets).
- First General Records Disposal Act passed by US Congress in 1889
- US Bureau of Efficiency Established 1912.
- US National Archives Founded 1934.
- First Records Disposition Schedule by the US National Archives in 1943

www.ArchiveBuilders.com 224004v027

C-119

History of Records Management (2)

- ARMA International (Association of Records Managers and Administrators) founded in 1956 as the American Records Management Association.
- ICRM (Institute of Certified Records Managers) founded in 1975. (Administers the CRM exam.)

www.ArchiveBuilders.com 224004v027

C-120

Document Management, Document Imaging

Retention Schedule

- Backup Plans Guarantee Data Integrity
- Retention Schedules Manage the Creation, Storage, and Destruction of Records
- Records Management and MIS Can Work Together to Integrate the Two Plans and Goals

www.ArchiveBuilders.com 224004v027

C-121

Manuals

- SGML or Die
- And Other Metadata Incantations
- 'Religious' Issues & Corporate Issues
- Electronic Signatures
- The Benefits of Fixed, Infrequent Publication Dates

www.ArchiveBuilders.com 224004v027

C-122

Updating Public Record Ordinances to Support Hyperlinked Objects on the Internet

www.ArchiveBuilders.com 224004v027

C-123

Public Record Ordinances

- Are Essential to Our Society and Our Freedom
- Require Submission on Paper
 - Discard Most Object Oriented Rules and Links
 - Disconnects Internet Document Hyperlinks
 - Erase the Fruits of Our Use of Technology
- Need To Be Updated
 - While Maintaining Their Goals In Support of Our Society
- [Spatial Indexing of Records]
 - Make all those hours with a joystick the right thing to do.
 - Bringing the mountain to Mohammed.

www.ArchiveBuilders.com 224004v027

C-124

On Organizing Documents

- Document Imaging does not magically organize unorganized records.
- Records should be organized before they are entered into a document imaging system.
- It is not wise to add records management problems to the imaging system implementation effort.

www.ArchiveBuilders.com 224004v027

C-125

Records Management Program

- Records Inventory
- Records Retention Schedule
 - All Types of Records
 - Origin
 - Physical Class
 - Function
 - Organizational Relationship
 - Applicable Regulations

www.ArchiveBuilders.com 224004v027

C-126

Document Management, Document Imaging

Lifecycles

www.ArchiveBuilders.com 224004v027

C-127

Forms Lifecycles

- Form Lifecycle
- Form Instance Lifecycle

www.ArchiveBuilders.com 224004v027

C-128

Records Life Cycle

- Creation
- Distribution and Immediate Use
- Storage and Maintenance
- Retention
- Disposition
- Archival Preservation

www.ArchiveBuilders.com 224004v027

C-129

Inactive Records

- Bankers Boxes (Records Cartons)
- Shelving
- Barcoding, Scan on Demand
- Inhouse
- Commercial Records Centers
 - 10 to 25 cents per month per box
 - 2 to 5 dollar retrieval fee per box

www.ArchiveBuilders.com 224004v027

C-130

Archives Management

- Appraisal
 - Review all Records
- Accessioning
 - File Archived Records in Archives
 - Prepare, Adapt Finding Aids for Use
- Protect and Preserve Forever

www.ArchiveBuilders.com 224004v027

C-131

Why Have Records Management? (1)

- Regulatory Compliance
- Business Operation
- Cost Containment
- Monitor New Technology
- Minimize Litigation Risk
- Safeguard Vital Information

www.ArchiveBuilders.com 224004v027

C-132

Document Management, Document Imaging

Why Have Records Management? (2)

- Control the Creation of Records
- Support Management Decisions and Planning
- Preserve the Corporate Memory
- Foster Professionalism in Business

www.ArchiveBuilders.com 224004v027

C-133

Records Management Barriers

- Records Management does not generate income.
- Records Management is not the organization's primary business.
- Most Records Management tasks are discretionary.

www.ArchiveBuilders.com 224004v027

C-134

Tasks Related to Records Management

- Forms Management
- Mail / Message Management / Internet Management
- Reprographics / Demand Printing / Report Distribution / COOL

www.ArchiveBuilders.com 224004v027

C-135

Summary

- Formats for Preservation in 3 Parts
 - **Meta data**, including indices, is required to interpret documents
 - **File format integrity**: to open a document file, you need the application, the OS, and the computer hardware (**Raster** format has the longest life.)
 - **Preserving the bits**: ECCs (Error Correcting Codes) recover bad bits. Copying restores ECCs.

www.ArchiveBuilders.com 224004v027

C-136

ISO 9000

- Quality Standard
- Uses Records to Support Quality Management
- Requires Explicit Procedures for Records and Information Management
- Manages Organization Wide Quality Including Sales and Marketing

www.ArchiveBuilders.com 224004v027

C-137

Do You Want to Be an ISO 9000 Library?

www.ArchiveBuilders.com 224004v027

C-138

Document Management, Document Imaging

Do You Want to Be an ISO 9000 University?

www.ArchiveBuilders.com 2240046027

C-139

Do You Want to Be an ISO 9000 Archive?

www.ArchiveBuilders.com 2240046027

C-140

Do You Want to Be an ISO 9000 Organization?

(Library, Archives, Museum, Records Center)
(Manufacturer, City, University, Law Firm)

www.ArchiveBuilders.com 2240046027

C-141

Document Imaging and Document Management for Litigation Support

www.ArchiveBuilders.com 2240046027

C-142

Document Processing

- Discovery, Production
- Scan, Identify, OCR, Store
- Code, Index, Workflow
- Import, Load, Merge, Link
- Display, Print, Transmit
- Search, Tag, Collaborate, Prepare Case
- In-Court Presentation
- Warehouse/Archive (To serve client interest)
- System Design, Sources of Information

www.ArchiveBuilders.com 2240046027

C-143

In-Court Presentation

www.ArchiveBuilders.com 2240046027

C-144

Document Management, Document Imaging

In-Court Presentation

- Three parts of vision
 - Resolution, detail (black and white)
 - Color
 - Motion
 - The foundation of optical tricks (illusions)
- Big, bright, high resolution
- Medium can carry the same weight as the presentation content

www.ArchiveBuilders.com 224004v027

C-145

Record Center Operation

www.ArchiveBuilders.com 224004v027

C-146

Is it Legal?

- Yes, it is analogous to microfilming.
- Anything can be done wrong.
- Records storage centers are a link in the physical chain of custody of records.
 - The chain can be broken through errors.
- Clever lawyers can create doubt in any situation.
 - Scanning is no different than any other business activity.

www.ArchiveBuilders.com 224004v027

C-147

Scan on Demand

- Competes with the cost of physical delivery
- Requires a computer system
 - Track scanned images
 - Provide security and passwords
 - Integration with records management system
- Must be managed
 - Or — it is just a fax system

www.ArchiveBuilders.com 224004v027

C-148

Indexing

- Box in warehouse
- Folder in box
- document in folder
- Name of document
- Document labels: date, number, client name, etc.
 - Full text will help avoid indexing
 - Client may do detailed indexing

www.ArchiveBuilders.com 224004v027

C-149

Fax Only

- If you have two databases of the same client information, the information for a given client will be different in the two databases.
- Scanning twice produces two copies.
 - The two copies will be different.
 - The two copies will be compared by the customer.
- Careful management of data reduces problems.

www.ArchiveBuilders.com 224004v027

C-150

Document Management, Document Imaging

Internet Delivery as an Email Attachment

- Modem (56 Kbits/s) : 3 pages per minute
 - (1 box per day)
- ISDN (128Kbits/s) : 10 pages per minute
 - 1 box in 1/2 day
- Cable (TV) Modem (500 Kbits/s): 1 page per second
 - 1 box per hour
- DSL (**D**igital **S**ubscriber **L**ine) (8 Mbits/s): 20 pages/s
 - 1 box in 2 minutes
 - Ranges from 1/2 to 8 Mbits per second depending on location.

www.ArchiveBuilders.com 224004v027

C-151

What Does the User Do With It

- Keep the scanned document as an email attachment.
- Add the document to a document management system.
 - Manually
 - Automatically
- Depend on you to find it for them when they need it again.

www.ArchiveBuilders.com 224004v027

C-152

Managing Imaging

- Not watching your imaging project closely is like leaving the wrong person in charge when you go to a conference or class.

www.ArchiveBuilders.com 224004v027

C-153

Politics

- An essential consideration
- Can we pretend it does not exist?
 - No
- Can we get past it with a quick maneuver?
 - No
- How do we deal with it?
 - Hard work every day

www.ArchiveBuilders.com 224004v027

C-154

RFP

www.ArchiveBuilders.com 224004v027

C-155

RFP

- RFP (*R*equest *f*or *P*roposal)
- RFI (*R*equest *f*or *I*nformation)
- Assorted Other Acronyms

www.ArchiveBuilders.com 224004v027

C-156

Document Management, Document Imaging

The Politics of an RFP from Authorship to System Operation

www.ArchiveBuilders.com 224004v027

C-157

The People Side

www.ArchiveBuilders.com 224004v027

C-158

What is the Magic Elixir, the Silver Bullet, the Secret Solution?

People

www.ArchiveBuilders.com 224004v027

C-159

The Contexts of an RFP

- Many different considerations affect an RFP.
- Each consideration lives in a world of its own.
 - And one world does not want to be concerned with another world.
- Writing an RFP is bringing harmony to these worlds.
 - Even if an RFP is not necessary, this is a good reason to write one.

www.ArchiveBuilders.com 224004v027

C-160

Context of an RFP

- People
- Technology
- Rush, Pressure
- City
- Time
- Politics
- Vendors
- Cost
- Other Cities
- State
- Federal
- Laws
- Knowledge Base
- Existing Computer Systems
- Existing Imaging Systems
- Interesting, But Not Burning Desire

www.ArchiveBuilders.com 224004v027

C-161

Personal RFP Contexts

- Everyone in this room has a context for an RFP.
 - Each person's context for an RFP is make up of a smidgen of each of the contexts listed.
- Everyone's context for an RFP is different.
 - Just as everyone's context for their organization is different.

www.ArchiveBuilders.com 224004v027

C-162

Document Management, Document Imaging

Talk the Talk, Walk the Walk

- Use the terminology well.
 - e.g. *Internet* or *Intranet*, as appropriate.
 - Put the right spin on your computer terms.
- Have many eyes check what you wrote.
 - Not just for legal and technical accuracy, but also for usage and style.
- Do everything well.
 - It has its own rewards.

www.ArchiveBuilders.com 224004v027

C-163

Logic

- Use logic to persuade:
 - It is logical to say “It is not logical to be logical.”
- If logic doesn’t work, use something else.
 - Fun, for example.
- Remember, document imaging creates digital analogs of the pages which we transmit over analog digital communications lines and make analog digital recordings of.
 - Not many Computer Scientists have contemplated digital analogs or analog digital signals.

www.ArchiveBuilders.com 224004v027

C-164

Silly, a Word in UNIX-Speak

- It describes a metaphor that pokes fun at being over-awed by the importance of system design.
- Translated to records management-speak, an example would be:
 - “Destroy these records, but make a file copy first.”
 - Or “Burn before reading.”

www.ArchiveBuilders.com 224004v027

C-165

Think the Unthinkable

- Could you cover it up?
 - Smaller is easier to cover up.
 - Have you made plans?
- Will everyone support you?
- How could you make the best of it?

www.ArchiveBuilders.com 224004v027

C-166

Vendors

- Vendor Considerations
 - Good Source of Information
 - Respect Their Profession
 - They Are On Your Team
 - It is Your Job to Provide Balance
 - It is Their Job to Sell (Use Their Selling Skills)

www.ArchiveBuilders.com 224004v027

C-167

Advice to Vendors

- Don’t Respond to an RFP Unless You Have Talked With the Person Who Wrote It.

www.ArchiveBuilders.com 224004v027

C-168

Document Management, Document Imaging

Should You Start with a Stock RFP?

- Yes, look at several.
- No, you must start from scratch, even if it looks a lot like one or more stock RFP's.
 - It must be your own work.
 - It is a lot safer that way.
- Some RFP's are very elaborate.
 - Write one that is appropriate to your organization.

www.ArchiveBuilders.com 224004v027

C-169

Sections of an RFP (1)

- Cover, Use Color
- Greetings Letter, Thank the Vendors for Taking the Time
- Organization Background, History
- Application, Today, in the Future
- Implementation Sequence

www.ArchiveBuilders.com 224004v027

C-170

Sections of an RFP (2)

- Budget Amount
- Payment Schedule
- Proposal Format
 - Suggested
 - Required
 - . . . and a Floppy
- Evaluation Criteria

www.ArchiveBuilders.com 224004v027

C-171

Sections of an RFP (3)

- Records Survey, Inventory
- Maintenance Costs
- Upgrade Costs and Path
- Training
- Existing Systems
- Vendor Background
- Insurance
- Legal

www.ArchiveBuilders.com 224004v027

C-172

RFP Response

- Let them know what you expect.
- Let them know who the contact(s) is.
- You can avoid fax responses.
- Let them put in brochures.
- Don't overdo the number of copies.

www.ArchiveBuilders.com 224004v027

C-173

Schedule of Events (RFP)

- Do you want a bidder's conference?
- Do you want a walk through of your site?
- Do you want demonstrations?
- Will you send out RFP changes? How? When?
- When do you want the responses?
- How will you extend the response deadline?

www.ArchiveBuilders.com 224004v027

C-174

Document Management, Document Imaging

Schedule of Events (System)

- What schedule will the system be installed, tested, accepted, turned over on?
- What is the payment schedule?
- What sort of project management will be used to track the project?
- When will you have meetings on the project?
 - Monthly?

www.ArchiveBuilders.com 224004v027

C-175

Vendor Background

- References
 - Customers
 - Trade
 - Financial
- History
- Resumes
 - Who will actually work on the project

www.ArchiveBuilders.com 224004v027

C-176

Legal

- Be sure legal signs off on it.
- Be sure you know what vendors think of it.
- Be sure that it controls vendors.
- Be sure that it does not control you.

www.ArchiveBuilders.com 224004v027

C-177

Legal Details (1)

- Specify a firm, fixed price.
- Make the proposal part of the contract.
- Use prior to acceptance does not constitute acceptance.
- Reserve the right to purchase the system, or any parts of the system, from any source.

www.ArchiveBuilders.com 224004v027

C-178

Legal Details (2)

- Require a prime contractor.
 - A single point of responsibility.
- Right to reject any and all proposals.
 - Right to reject any or all parts of any proposal.
- Unavailability of funds:
 - If funds are unavailable, reserve the right to postpone or cancel the award.
 - Reserve the right for convenience.

www.ArchiveBuilders.com 224004v027

C-179

Legal Details (3)

- Specify how to take exception to the RFP.
- Specify a method of withdrawing proposals.
- Specify rights to pertinent material.
 - All materials become the property of the city.
 - Confidentiality of documents.
 - Non-disclosure.

www.ArchiveBuilders.com 224004v027

C-180

Document Management, Document Imaging

Legal Details (4)

- Cost of proposal not paid by the city.
 - Consider funding a study.
- Specify how you will modify the RFP.
- Restrictions on contact with city employees.
- Other general conditions.
 - See your legal department.
 - Make sure they sign on.
- For each detail, vendors have a response.

www.ArchiveBuilders.com 224004v027

C-181

System Components

- Available Off-the-Shelf
- Of Current Manufacture
- In Regular Production
- New (not used)
- Not Beta
- Not A Prototype
- Not a Special
- Guaranteed Support in All Future Releases

www.ArchiveBuilders.com 224004v027

C-182

Features & Benefits

- Request that vendor responses categorize each listed feature as:
 - Available in Base System
 - Can Be Added at No Cost
 - An Extra Cost Item
 - Not Available
 - Negotiable
 - Other

www.ArchiveBuilders.com 224004v027

C-183

Response Time

- A second or two for images
 - More for jukeboxes (autochangers)
- Milliseconds for database inquiries
- Under what type of system load
- Allow yourself a way to relax the requirement.
- Prevent vendors from unilaterally relaxing the requirement.

www.ArchiveBuilders.com 224004v027

C-184

Source Code

- What if the software vendor disappears?
- For big vendors, the users band together.
- For small vendors, software can be put in escrow.
- You will eventually have to export your documents to the system of a surviving vendor.

www.ArchiveBuilders.com 224004v027

C-185

Theft

- Due diligence
 - References
 - Contacts
 - Years of familiarity with industry players
- Some legal protection (words)
- You must know the people you are dealing with.

www.ArchiveBuilders.com 224004v027

C-186

Document Management, Document Imaging

Records Survey

- Based on your records management system.
- Include yearly growth.
- Provide schedule for backfile conversion to document imaging.
- Include document size, binding, and quality.
- You should consider including some or all electronic records.

www.ArchiveBuilders.com 224004v027

C-187

RFP Sizes

- Big
 - Cost as much as a small imaging system
- Small
 - Less Risk
 - Less controls required

www.ArchiveBuilders.com 224004v027

C-188

Checks and Balances

- Big RFP:
 - Can be written by a consulting firm.
 - You could use a small outside party to watch the two big outside parties.
 - the RFP consulting firm
 - the system vendor
- Small RFP:
 - You could use a small outside party to check on one not-as-big outside party. (The system vendor.)

www.ArchiveBuilders.com 224004v027

C-189

RFP's for Thespians

www.ArchiveBuilders.com 224004v027

C-190

First, Someone Gets a Bee in Their Bonnet

- Hopefully they immediately come to you.
- Soon there is a committee.
- A camel is a horse designed by a committee.

www.ArchiveBuilders.com 224004v027

C-191

The first Reason for an RFP Is to Let the Vendors Do All the Work

- Throw in whatever anyone can think of, ask all your questions with a 'must' or 'will'.
- Vendors want some return on their effort.
- No one is sure if they really want a system.
- Sometimes something is actually purchased, usually providing tar and feathers for all involved.

www.ArchiveBuilders.com 224004v027

C-192

Document Management, Document Imaging

Under Ideal Circumstances, You Do Your Homework

- You design a system that meets your needs.
- Most components in your design are from specific vendors.
- You could live with other vendors for some of the components.
- You have your design tested by experts.

www.ArchiveBuilders.com 224004v027

C-193

You Write an RFP for the System

- Not everyone gets on board with you, the strays are plotting your demise.
- You did not make sure that legal said everything they wanted to.
- The vendors that responded to last year's RFP are angry that you do not favor them.
- Disgruntled vendors and last year's system champions are working with the people on the sidelines to delay, suspend, supersede, or terminate your RFP.

www.ArchiveBuilders.com 224004v027

C-194

Changes (1)

- You did not talk legal out of requiring something that your preferred vendor can not do.
- You did not realize that some of the component specifications had subtle meanings, knocking out your preferred vendors.

www.ArchiveBuilders.com 224004v027

C-195

Changes (2)

- The purchasing department added a few specifications from a vendor they knew, creating a lock-out-spec for that vendor's products.
- You forgot to put in the error disclaimer. You made several last minute changes to correct problems related to last minute technical discoveries. The changes introduced more errors than you corrected.

www.ArchiveBuilders.com 224004v027

C-196

You Issue the RFP (1)

- A spectacular new product is introduced. You forgot to leave yourself a method to amend your RFP.
- Your preferred vendor is five minutes late in delivering their bid. You forgot to add a method to extend the RFP deadline.
- You like parts from more than one vendor. Your RFP did not allow you to mix and match.

www.ArchiveBuilders.com 224004v027

C-197

You Issue the RFP (2)

- You did not put in the 'legal' to deny the cost of responding to the RFP. Three vendors bill you and go to court.
- You do not like any of the responses. You forgot to put in the 'legal' to cover 'no award'.
- You forgot your non-disclosure agreement. Your competitors now know how you do business to a very fine level of detail.

www.ArchiveBuilders.com 224004v027

C-198

Document Management, Document Imaging

You Evaluate the RFP Responses (1)

- Your RFP response evaluation committee excludes you from their meetings.
- The evaluation committee assigns the wrong values or scores to the evaluation criteria you selected.

www.ArchiveBuilders.com 224004v027

C-199

You Evaluate the RFP Responses (2)

- Someone changes the evaluation criteria and two vendors sue.
- You are told to manage the evaluation scoring so that a newly acquired division will win the contract.

www.ArchiveBuilders.com 224004v027

C-200

A Vendor Wins the RFP (1)

- The winning vendor does not meet the specifications and you cannot explain, nor can your committee understand, the differences between the vendor's specifications and the RFP requirements.
- (Ditto for differences in the specifications between what is delivered and what was specified by the vendor.)

www.ArchiveBuilders.com 224004v027

C-201

A Vendor Wins the RFP (2)

- You realize that this project will cost more than any other project ever undertaken by your organization.
- You remember someone saying that this was a "You bet your job" deal.

www.ArchiveBuilders.com 224004v027

C-202

Installation Starts (1)

- All time and materials work requires \$300 dollar per hour people from your vendors headquarters.
- Everything becomes time and materials.
- Your vendor moves, merges, is purchased, goes out of business. Twice.

www.ArchiveBuilders.com 224004v027

C-203

Installation Starts (2)

- The documents you were going to image:
 - Increase by 300 percent
 - Disappear in a merger
 - Are changed to light blue NCR (No Carbon Required) paper with light blue type
 - Are replaced by documents that come in 57 formats and 16 sizes.

www.ArchiveBuilders.com 224004v027

C-204

Document Management, Document Imaging

Alternatives to an RFP

www.ArchiveBuilders.com 224004v027

C-205

Phased Approach: Under the Bar (1)

- Know what dollar amount requires an RFP.
- Small expenditures are easier to explain if something goes wrong.
- Mid-course corrections are hard.
- Get finished often so you can see where you are -- often.

www.ArchiveBuilders.com 224004v027

C-206

Phased Approach: Under the Bar (2)

- Every installation is a learning experience.
- The faster you can get done, and apply the learning to the next phase, the better.

www.ArchiveBuilders.com 224004v027

C-207

Planning (1)

- A phased approach depends on good planning.
- Use what the vendor has. Force is expensive, let the market place do it. Wait until next year.

www.ArchiveBuilders.com 224004v027

C-208

Planning (2)

- Careful planning and a good understanding of the technology is required so that the little pieces build a system and not just a collection of little pieces.
- Vendors are pricing alternatives with pilot systems in mind.

www.ArchiveBuilders.com 224004v027

C-209

Paradigms & References

www.ArchiveBuilders.com 224004v027

C-210

Document Management, Document Imaging

Paradigms

- Examples:
 - IT should have more money
 - Cost centers should economize
 - I just know. It just is. Everyone knows.
 - Nothing outside my paradigm exists (makes sense)
 - A person's paradigm encompasses everything
 - Traditional: There is just one paradigm
 - So, why mention it?
 - It is not just my paradigm, it is everyone's
 - I am right, you are wrong

www.ArchiveBuilders.com 224004v027

C-211

Paradigms

- An idea from a different paradigm is like a monster that appears from out of nowhere, ready to eat you.
 - This is particularly true if your paradigm is that there is only one paradigm, so there is no need to discuss it.

www.ArchiveBuilders.com 224004v027

C-212

Recontextualization

- By definition, because a paradigm defines the entire universe, it already has a context for everything.

www.ArchiveBuilders.com 224004v027

C-213

Blind Spot

- The way we fit something outside of our paradigm into our paradigm.
- A teacher who had just studied the nine dots 'outside the box solution' presented it to their class.
 - A student folded the paper with the nine dots into pleats and connected all nine dots with a single stroke
 - The teacher marked the solution incorrect

www.ArchiveBuilders.com 224004v027

C-214

Paradigm: Division and Grouping: the Scientific Method

- Why don't shellfish have fins?
 - Because we said that everything in the ocean was a fish.

www.ArchiveBuilders.com 224004v027

C-215

Paradigms

- New: paradigms of paradigms
 - A basic requisite to support diversity
 - Each person has their own unique paradigm
 - Each person has an equally valid paradigm
 - A source of equality
 - Who's paradigm should we use for validation?
 - Paradigms are self validating
 - Win-win is only possible in a third paradigm
 - Compromise is lose-lose in a single paradigm
- The need for a passionate project champion is not in all paradigms

www.ArchiveBuilders.com 224004v027

C-216

Document Management, Document Imaging

Philosophy: Paradigms of Paradigms

- Everyone has a unique paradigm
 - No two persons' paradigms are alike
 - Each person accommodates the paradigms of all other persons
- Each person's paradigm defines the world
 - If a person has a non-technical paradigm, then technology must fit in a non-technical world
 - If a person has a technical paradigm, then technology explains many aspects of the world

www.ArchiveBuilders.com 224004v027

C-217

What are We Preserving? How are We Preserving It?

- Wisdom
- Knowledge
- Information
- Data
- Bits
- Discernable differences

www.ArchiveBuilders.com 224004v027

C-218

Moving Up A Level

- More is not the answer
 - No matter how discernable, a difference does not make a bit.
 - There is no number of bits that will produce data
 - More data does not produce information
 - More information does not produce knowledge
 - More knowledge does not produce wisdom

www.ArchiveBuilders.com 224004v027

C-219

Moving Up: How is it Done?

- Computers can not do it.
 - If they could, they would be people.
 - Moving up lies in creating the next level.
- We can do it.
 - But we do not know how.
- We can teach it.
 - But, we do not how we teach it.

www.ArchiveBuilders.com 224004v027

C-220

Serving Up a Bucket of Wisdom

- Wisdom does not come in buckets.
- Wisdom does not come at all.
- People can act wisely.
- You have acted wisely in the past.
- You can act wisely now.
- Questions are more important than answers.
- We can talk about and teach wisdom.
- Most things are easier than wisdom.
 - Wisdom is a good choice.

www.ArchiveBuilders.com 224004v027

C-221

Creation of a Library

- What is a library? How long does it take to create, or to destroy, a library?
 - “We will have a library. Lock the books room. Write procedures. Restrict and manage access.”
 - “We do not need a library. Unlock the books room. Throw away the procedures. Let anyone who needs the books have them.”
- Both are ideas. Both occur in an instant.

www.ArchiveBuilders.com 224004v027

C-222

Document Management, Document Imaging

For More Information . . .

www.ArchiveBuilders.com 224004v027

C-223

For More Information . . .

- Certifications
- Organizations
- Magazines
- Websites
- Shows
- Catalogs

www.ArchiveBuilders.com 224004v027

C-224

Certifications

- CRM (**C**ertified **R**ecords **M**anager)
- CDIA (**C**ertified **D**ocument **I**maging **A**rchitect)
- CNE (**C**ertified **N**etwork **E**ngineer) (Novell)
- MSxx (**M**icrosoft **C**ertified **P**rofessional)
- MIT (AIIM **M**aster of **I**nformation **T**echnology)
- LIT (AIIM **L**aureate of **I**nformation **T**echnology)
- CA (**C**ertified **A**rchivist)

www.ArchiveBuilders.com 224004v027

C-225

Organizations (1)

- AIIM (*Association for Information and Image Management*)
– +1 (800) 477-2446 +1 (301) 587-8202
– <http://www.AIIM.org>
- ARMA (*Association of Records Managers and Administrators*)
– +1 (800) 422-2762 +1 (913) 341-3808
– <http://www.ARMA.org>

www.ArchiveBuilders.com 224004v027

C-226

Organizations (2)

- PRISM International (*Professional Records and Information Services Management International*) Formerly ACRC (*Association of Commercial Records Centers*) +1 (800) 336-9793 +1 (919) 881-0677 <http://www.PRISMIntl.org>
- SAA *Society of American Archivists*
- SCA *Society of California Archivists*

www.ArchiveBuilders.com 224004v027

C-227

Organizations (3)

- LRMA (*Legal Records Management Association*)
- SLA (*Special Libraries Association*)
- ASIS (*American Society for Information Science*)
- ACP (*Association of Contingency Planners*)
- ASQ (*American Society for Quality (ISO 9000)*)
- Xplor (*Print on Demand*)
- GCA (*Graphics Communications Association (Printing)*) +1 (703) 519-8160
– SGML (*Structured Generalized Markup Language*)

www.ArchiveBuilders.com 224004v027

C-228

Document Management, Document Imaging

Magazines (1)

- Adobe Magazine +1 (206) 628-2321
- Advanced Imaging +1 (516) 845-2700
- Computer Telephony +1 (215) 355-2886
- Document Management +1 (602) 585-5580
- Government Information and Imaging Technology +1 (301) 445-4405
- Imaging +1 (800) 999-0345, +1 (800) 542-7279

www.ArchiveBuilders.com 2240046027

C-229

Magazines (2)

- Imaging Business (Resellers) +1 (301) 340-1520
- Imaging World +1 (207) 236-8524
- Inform (AIIM) +1 (800) 477-2446
- Law Technology +1 (800) 888-8300
- PC Week +1 (609) 786-8230 (A must for computing)

www.ArchiveBuilders.com 2240046027

C-230

Magazines (3)

- Print on Demand Business +1 (212) 961-1906
- Southern California Micro Publishing News +1 (310) 371-5787
- Windows NT Magazine +1 (800) 621-1544
- Back Office Magazine +1 (918) 832-9252

www.ArchiveBuilders.com 2240046027

C-231

(Hint)

- Use magazines as a vocabulary test
– . . . And as a vocabulary builder.

www.ArchiveBuilders.com 2240046027

C-232

Shows

- AIIM, Early April (Biggest)
- Comdex Fall, November, Week before Thanksgiving, Las Vegas (every year)
- Computer Telephony, March, Los Angeles Convention Center +1 (215) 355-2886
- Documation (SGML) (GCA) March
- Legal Tech, June, Los Angeles, (Good Show) +1 (212) 877-5619

www.ArchiveBuilders.com 2240046027

C-233

How do I Keep Up? - Websites

- Microsoft Press Release Page
– <http://www.Microsoft.com/presspass/todaynews.htm>
 - This is a good page to use as your Internet start page
- PC Week
– <http://www.ZDNet.com/PCWeek/>
 - e.g. — Internet2 will link 150 educational institutions at OC 192 (Optical Carrier) (10 Gigabits per second) starting February 24, 1999 using SONET (Synchronous Optical Network) links donated by Qwest Communications

www.ArchiveBuilders.com 2240046027

C-234

Document Management, Document Imaging

Catalogs

- Light Impressions (Archival Supplies)
+1 (800) 828-6216
- See Also: Guides for Shows

www.ArchiveBuilders.com 224004v027

C-235

Books

www.ArchiveBuilders.com 224004v027

C-236

Books as Computer Printouts of Library, Archival, and Museum Materials

- Treasures of the Library of Congress by Charles A. Goodrum / Published 1991
- The National Archives of the United States by Herman J. Bicla / Published 1984
- The Smithsonian : 150 Years of Adventure, Discovery, and Wonder by James Conaway / Hardcover / Published 1995
- Treasures of the Smithsonian by Edwards Park / Published 1994

www.ArchiveBuilders.com 224004v027

C-237

This is a Good Records Management Book To Have on Your Book Shelf

- Even if you don't read it.
- Information and Records Management
– By Mary F. Robek, et al.
 - Published by Glencoe
 - ISBN 0-02-8017923-5
 - Available from ARMA +1 (800) 422-2762
- So that when you need it, you will have it.

www.ArchiveBuilders.com 224004v027

C-238

XML

- The Xml Handbook (Charles F. Goldfarb Series) by Charles F. Goldfarb, Paul Prescod

Unicode

- The Unicode Consortium. *The Unicode Standard, Version 3.0*
Reading, MA, Addison-Wesley Developers Press, 2000. ISBN 0-201-61633-5.

www.ArchiveBuilders.com 224004v027

C-239

Adobe

- Adobe Systems, *PostScript Language Reference*, Third Edition, Addison-Wesley, Menlo Park, CA, 1999, ISBN 0-201-37922-8
<http://www.adobe.com/products/postscript/pdfs/PLRM.pdf>
- Adobe Systems, *PostScript Language Reference Supplement*, Adobe PostScript 3, Version 3010 and 3011, Product Supplement, 1999, 28-007
<http://partners.adobe.com/asn/developer/PDFS/TN/PS3010and3011.Supplement.pdf>
- Adobe Systems, Inc., *Adobe Type 1 Font Format*, Addison-Wesley, Menlo Park, CA, 1990 ISBN 0-201-57044-0
http://partners.adobe.com/asn/developer/PDFS/TN/T1_SPEC.PDF
- Adobe Systems, *Type 1 Font Format, Supplement*, 1994, Technical Specification #5015, PN (Part Number) LPS5015
http://partners.adobe.com/asn/developer/PDFS/TN/5015.Type1_Supp.pdf

www.ArchiveBuilders.com 224004v027

C-240

Document Management, Document Imaging

Adobe

- Adobe Systems, *Portable Document Format Reference Manual*, Version 1.3, Addison-Wesley, Menlo Park, CA, 1999, ISBN 0-201-62628-4
<http://partners.adobe.com/asn/developer/acrosdk/DOCS/pdfssec.pdf>
- Adobe Systems, *Display Postscript Manual*, 1993
<http://partners.adobe.com/asn/developer/PDFS/TN/DPS.refmanuals.PSW.pdf>
- Adobe Extreme, brochure (integrates PDF and Postscript Printing), 1998, Adobe Systems <http://www.adobe.com/products/extreme/pdfs/extremewp.pdf>
- Ried, Glenn C., *PostScript Language Program Design*, Addison-Wesley, Menlo Park, CA, ISBN 0-201-14396-8 1988
- Adobe Systems, *PostScript Language Tutorial and Cookbook* Addison-Wesley Menlo Park, CA, 1986 ISBN 0-201-10189-0
- Adobe Systems, *PostScript Language Document Structuring Conventions Specification*, Version 3.0, Adobe Systems, 1992, PN LPS5001
http://partners.adobe.com/asn/developer/PDFS/TN/5001.DSC_Spec.pdf

www.ArchiveBuilders.com 224004v027

C-241

The Best Bill Gates Book I have Read

- Gates: How Microsoft's Mogul Reinvented an Industry - and Made Himself the Richest Man in America

– By Stephen Manes and Paul Andrews

- Simon & Schuster 1993
- ISBN 0-671-88074-8

www.ArchiveBuilders.com 224004v027

C-242

Since the Book was Published

- Bill has quadrupled his fortune.
- Bill is about to add a zero.
 - (CentiBillionaire, USD)
- He may add another zero in the next ten years.
 - His nickname was once Trey.
- Bill tithed in 1999. US\$ 17 Billion.

www.ArchiveBuilders.com 224004v027

C-243

The Best NT Social History (Unauthorized) I have Read

- Showstopper, The Breakneck Race to Create Windows NT and the Next Generation at Microsoft

– By G. Pascal Zachary

- The Free Press (Div. of Macmillan, Inc.)
- ISBN 0-02-935671-7

www.ArchiveBuilders.com 224004v027

C-244

The Best NT Social History (Authorized) I have Read

- Inside Windows NT
 - By Helen Custer
 - Foreword by David N. Cutler
 - Microsoft Press
 - ISBN 1-55615-481-X

www.ArchiveBuilders.com 224004v027

C-245

Change

www.ArchiveBuilders.com 224004v027

C-246

Document Management, Document Imaging

Too Fast is not Fast Enough

- Change in Our Future
- If you say something unbelievable, people will not believe you.

www.ArchiveBuilders.com 224004v027

C-247

Change: Learning to Use the Internet

- Even the people who make the most use of the Internet only use a part of its capabilities.
- These power users make up far less than one percent of the population.
- Huge changes would occur if the lives of even ten percent of the population became as Internet-centric as the lives of these power users.
- But, as great, and as fast, as these changes are sure to be, there are bigger and faster changes coming.

www.ArchiveBuilders.com 224004v027

C-248

The Internet Will Change Faster Than We Can Learn to Use It

- Because, even as we learn to use it, the Internet itself is changing.
- Terabit per second fiber optic transmitters and receivers will be introduced by Nortel Networks in 2001
- In five years, DVD quality video-telephony could be as free and as available as email is today.
 - If this seems impossible, remember that five years ago, free worldwide email seemed impossible.
-

www.ArchiveBuilders.com 224004v027

C-249

Displacements

- With free video-telephony, the jobs of receptionists in offices may be exported to locations around the world.
- Video stores will likely disappear.
- Tractor drivers may work from other countries.
 - The 95 percent drop in rural populations may occur again.
- A simple change in the quality-of-service Internet protocol, a programming change, can provide free CD quality telephony worldwide with existing Internet hardware.
 - The existing telephony infrastructure will become redundant

www.ArchiveBuilders.com 224004v027

C-250

Tools and Responsibilities

- These are just some of the changes that Information Resource Managers, and all other professionals, will make use of in their jobs.
- The information economy will continue to grow, creating jobs for everyone skilled in information management.
- The world will look to these professionals for the benefits that their work will bring
- More importantly, the world will look to these professionals for the responsible husbanding of our most fundamental social values.

www.ArchiveBuilders.com 224004v027

C-251

Nanotechnology

www.ArchiveBuilders.com 224004v027

C-252

Document Management, Document Imaging

Nanotechnology: 15 Years Old

- Getting massive funding in the US Federal budget
 - Which has been expanded to a USD \$500 million US National Nanotechnology Initiative.
- Physical objects become just printouts of electronic documents
 - Like the food replicators on Star Trek.
 - Information and objects become one in the same.
 - Only organisms that have memories and environmentally affected physical adaptations will not be identical with information
 - PCs build up injuries of bad system files and cruft on their hard drive
 - We have grown up with it, so it is transparent, like water to a fish
 - Work-arounds are healed wounds.
 - The premise of the movie Matrix will become more possible.

www.ArchiveBuilders.com 224004v027

C-253

Nanotechnology

- Is based on molecular machines.
 - <http://www.foresight.org/>
- Will replace computers by the year 2050.
 - And all of technology as we know it.
 - Mechanical nano-computers are 1 million times faster than the 1 GigaHertz computer we have now.
- Example (Juan Ponce De Leon (1460?-1521)):
 - In vivo correction of all DNA errors that build up over time in all cells.

www.ArchiveBuilders.com 224004v027

C-254

Nanotechnology

- The equations were completed in the 1980s
- They are self replicating
 - They are free
 - It takes one to make one and we do not have any

www.ArchiveBuilders.com 224004v027

C-255

The Fear of Key-Lime Pie Filling

- An off-switch for every organism, certified organic will have considerably more meaning
- The end of the green house effect
- Accidental cessation of evaporation by a 12 year old who's parents were born last year in Victoria, British Columbia, Canada
 - In a 'Get Smart' episode, Maxwell Smart had a pen that would cause a chain reaction, converting all water to a new state of ice.
 - There was not complete certainty that the first atomic bomb (or the first hydrogen bomb) detonation would stop.
- Key-lime pie filling, an industrial accident
- One of the few things that will still be of interest will be the records we have kept.

www.ArchiveBuilders.com 224004v027

C-256

Vision and Systems Success

- Microsoft is Bill Gates' vision.
- Nanotechnology is Eric Drexler's vision.
- Your Library, Archives, Museum, Records Center, and document management systems will spring forth from your vision.

www.ArchiveBuilders.com 224004v027

C-257

Thank You

www.ArchiveBuilders.com 224004v027

C-258