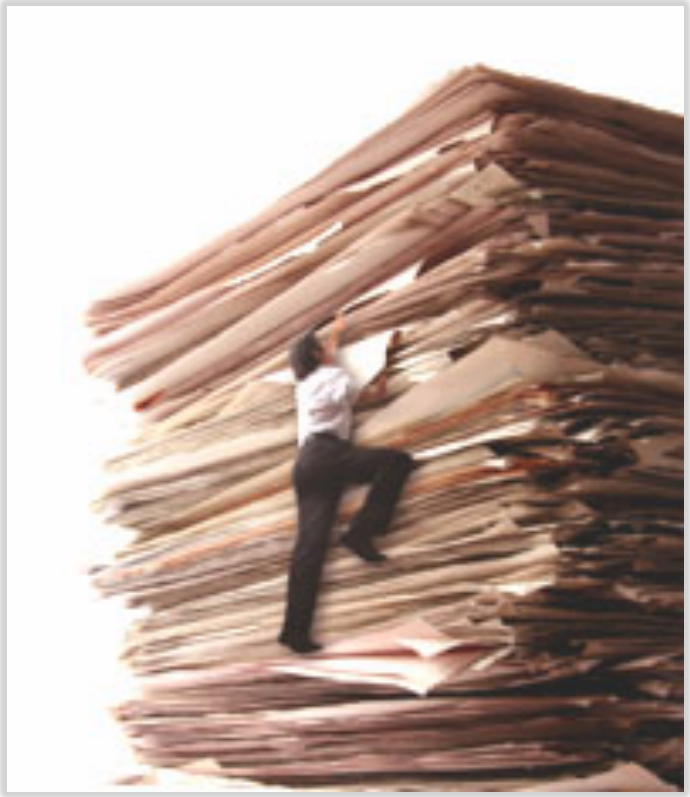
THE OHIO STATE UNIVERSITY

ELECTRONIC RECORDS PRESERVATION

NO RECORDS MANAGEMENT?



25%

NO RECORDS MANAGEMENT?

Every 12 seconds 1 document is lost.

28,800 seconds (8 hour day)/12

=2400 documents lost each day

No Records Management?

Computer users spend 7.5% of their time on a computer looking for files.

260 work days/year

2080 work hours/year

x 0.075

= 156 hours/year

= 19.5 days/year

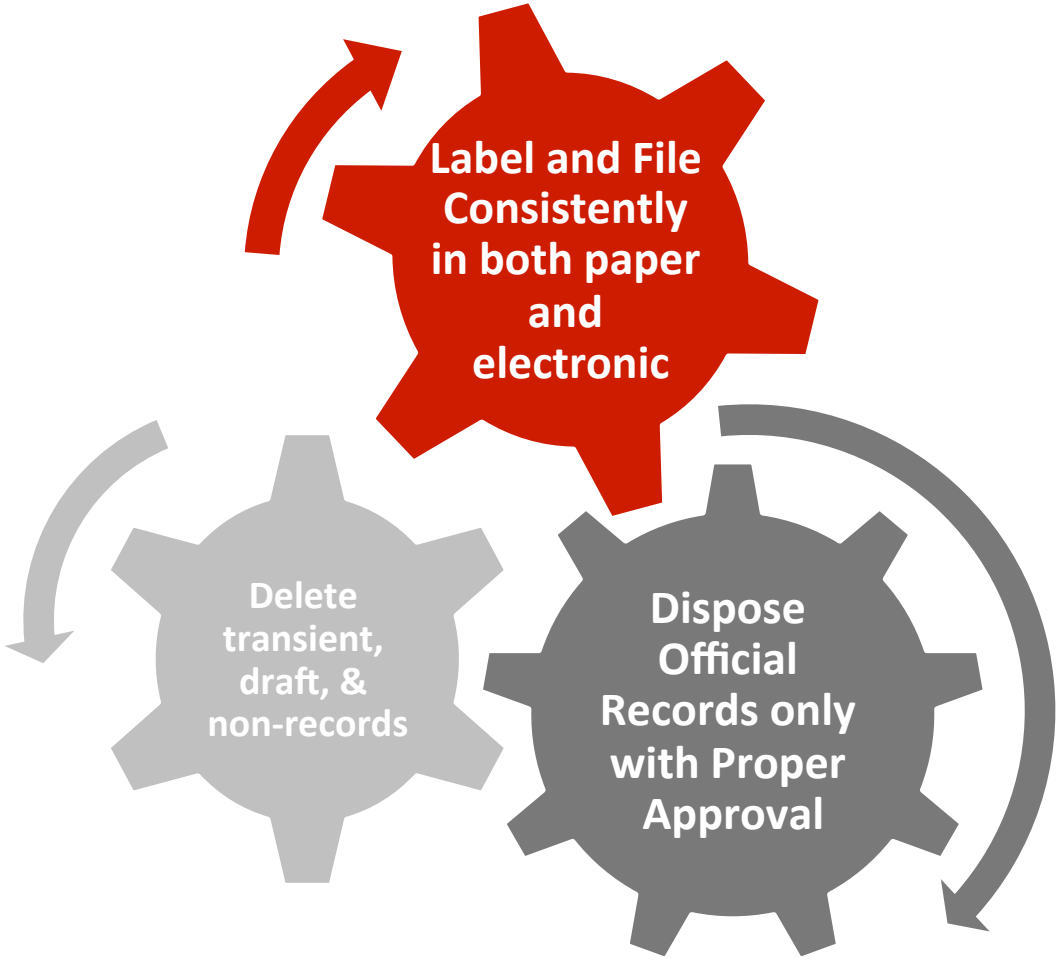
Time spent looking for files.

Purpose of Records Management

- Ensure that records are available when needed
  - Internally for office functions
  - Public records requests
  - Discovery
- Protect records from improper or unauthorized destruction
- Ensure that records are not retained unnecessarily
  - Destroy records when retention period has expired

Retaining records too long can be as much of a liability as not retaining them for long enough.

EVERYONE'S RESPONSIBILITIES



GOAL OF A RECORDS MANAGEMENT PROGRAM

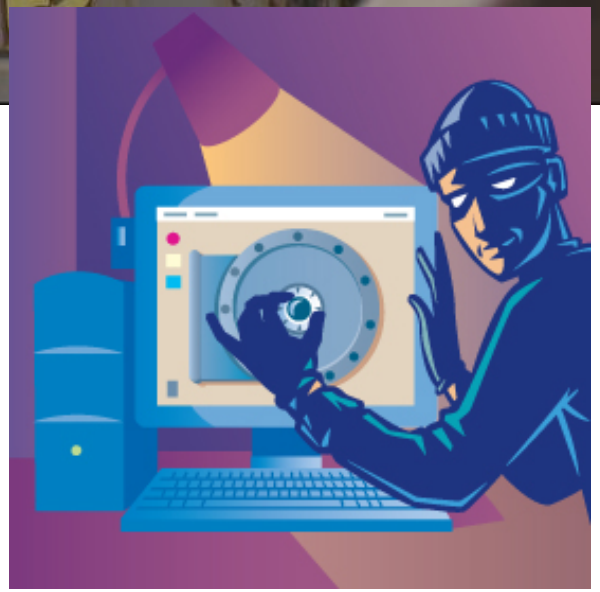
Establish a ***systematic***, ***repeatable***, ***documented*** ***process*** for the retention and disposition of records created and maintained by the public office.

Benefits of Records Management

TRANSPARENCY IN GOVERNMENT

Public's Records – Not Yours!

DECREASED RISK OF LITIGATION



Removal of Public Records

All records are the property of the public office concerned and shall not be removed, destroyed, mutilated, transferred, or otherwise damaged or disposed of, in whole or in part, except as provided by law or under the rules adopted by the records commission.

---ORC 149.351

BENEFITS OF RECORDS MANAGEMENT

Transparency in Government

Decrease Risk in Litigation

Increased Efficiency

Save Resources

Records Management in Policy

Repeating Concepts

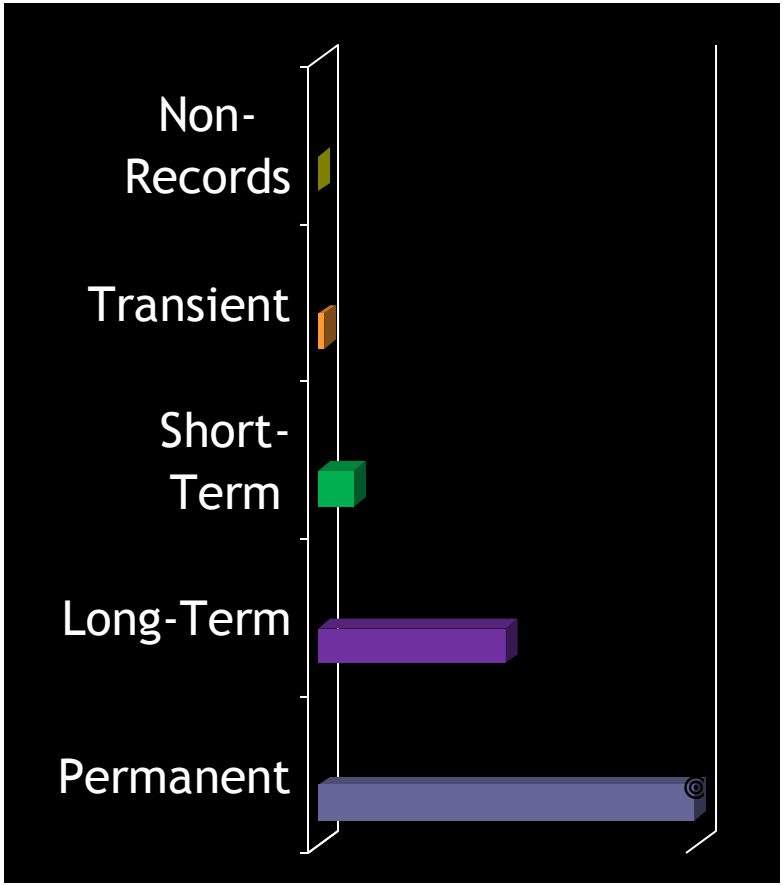
- Retention based on content
- Retain in accordance with applicable retention schedule(s)
- Employees are responsible

Across Institutional Policies

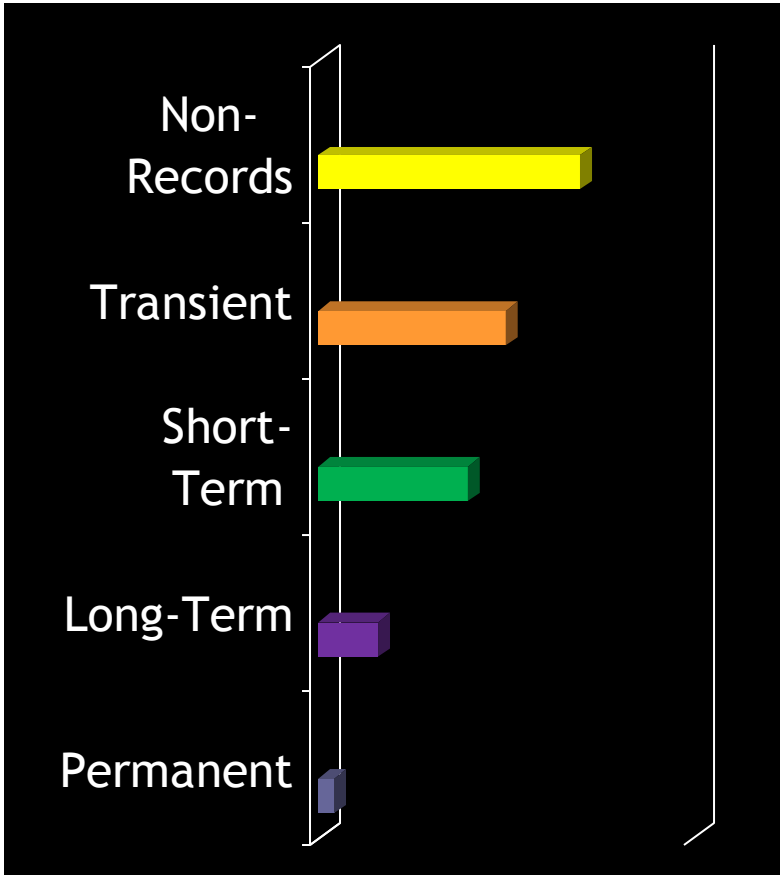
ORGANIZING ELECTRONIC RECORDS

RETENTION V. QUANTITY

Retention Time



Quantity



The Pacman® Effect…

Here to There

- Save it all “just in case”
- Ownership “my email”
- Lost productivity searching for information

- Defensible deletion with retention schedules
- Reduced impact of staff changes
- Reduced information retrieval time

Dairy | Eggs | Milk | Meat | Paper Goods Grill/Seasonal

Dairy

Deli

Deli/ Dairy

Wine / Beer / Tea & Coffee

Cookies, Snacks, Soda Pop

Snacks and Boxed Juice

Baking, Baking Supplies, Spices

Pasta, Sauce, Canned Fruit

Rice, Boxed Dinners, Condiments

Cereal, Jams, Spreads

Frozen Dinners & Meals

Frozen Vegetables & Desserts

Cards, Toy, home office, & sports goods

Seasonal, Candy

Candy, Seasonal, Air Fresheners

Pallets - promotional

Paper Goods

Trash Bags, Food Storage, Cleaners

Laundry Care, Floor, Dish Cleaner

Cat & Pet

Dog

Baby

Cosmetics / Lotions

Hair Care. Deodorant

Oral Care. Shaving

Bath / OTC

OTC / Nutrition

Vitamins / Nutrition

Eye & Foot Care, Feminine Care

Hair , Apparel, Basic DIY

Pharmacy

Customer service

7 check-outs + 3 self check-outs

Bakery

Produce

Fresh Produce

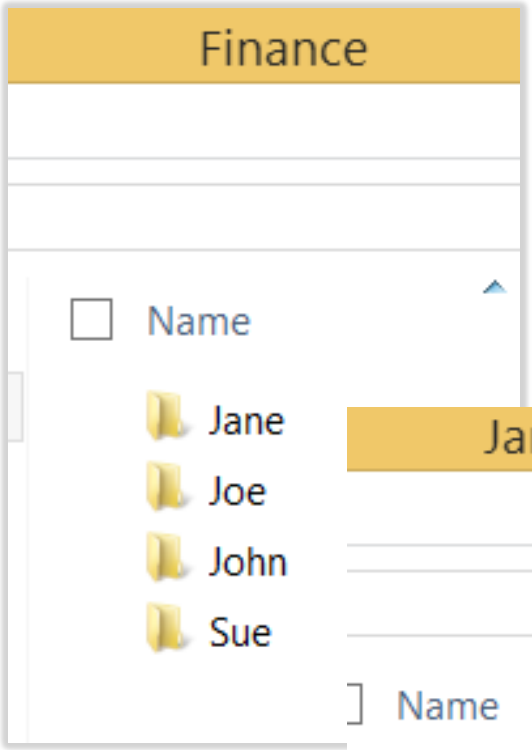
Enter Exit

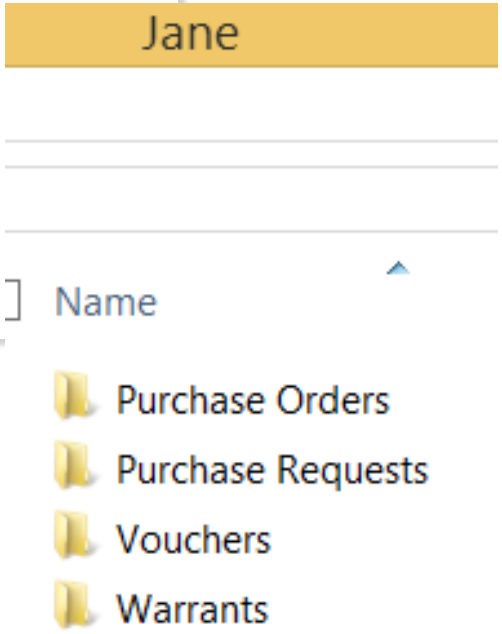
Photo Processing

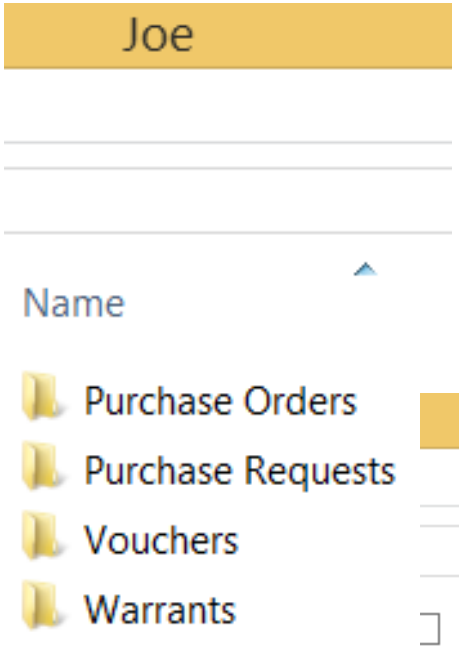
Unmanaged Shared Drives

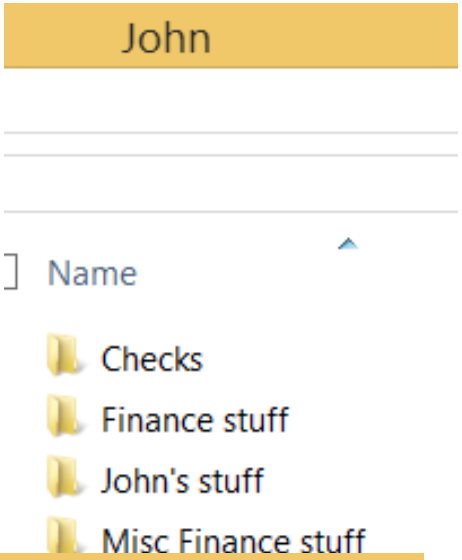
- Can contain important documents & records
  - But is intermingled with useless information
    - Duplicates
    - Non-Records
    - Records past retention time
- Generally lacks standard naming conventions
- Ownership is unclear
  - Terminated/transferred employees
    - Nobody wants to take responsibility for remaining records

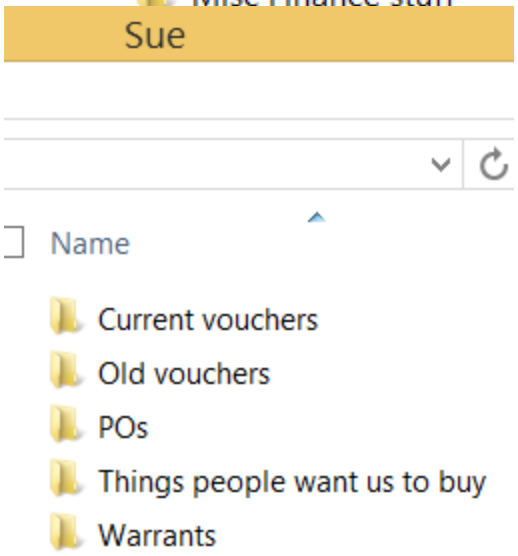
Not Ideal


Finance
Name
Jane
Joe
John
Sue


Jane
Name
Purchase Orders
Purchase Requests
Vouchers
Warrants

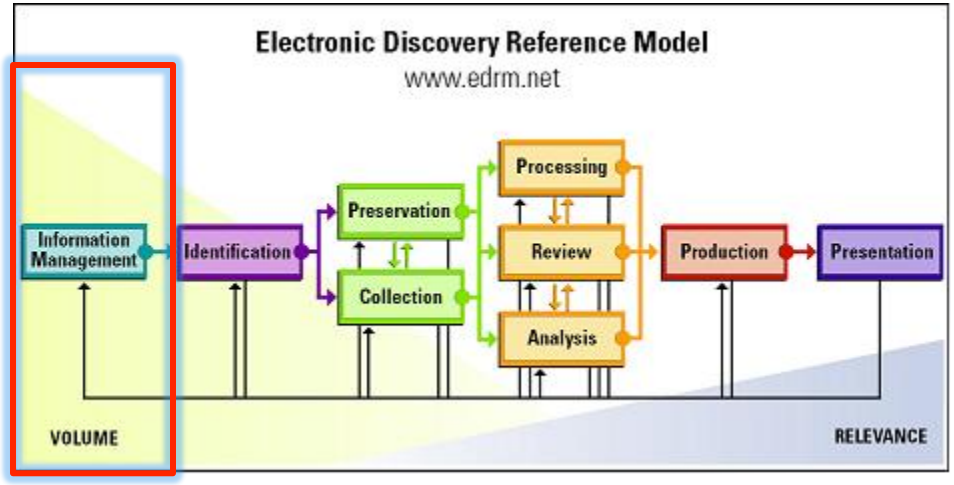

Joe
Name
Purchase Orders
Purchase Requests
Vouchers
Warrants


John
Name
Checks
Finance stuff
John's stuff
Misc Finance stuff

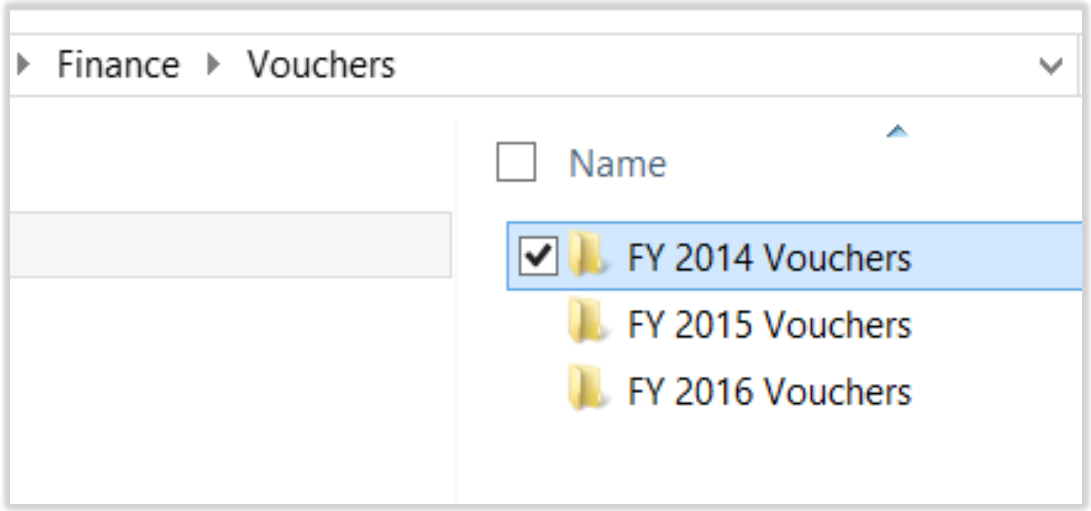

Sue
Name
Current vouchers
Old vouchers
POs
Things people want us to buy
Warrants

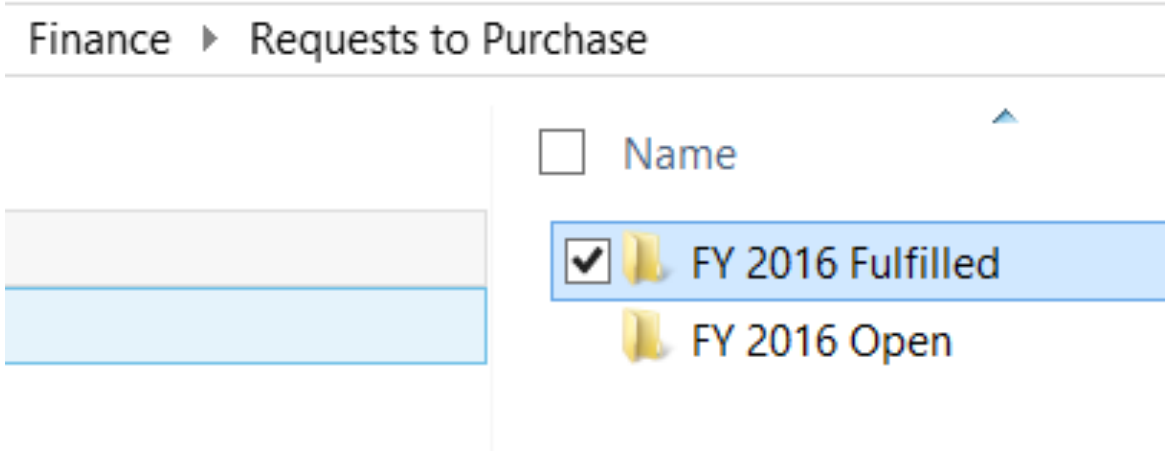
Transforming Network Shared Drives

- Cleaning, classifying, and organizing content on network shared drives to improve information management
  - Ability to locate
  - Defensibility
  - Storage
  - Retention

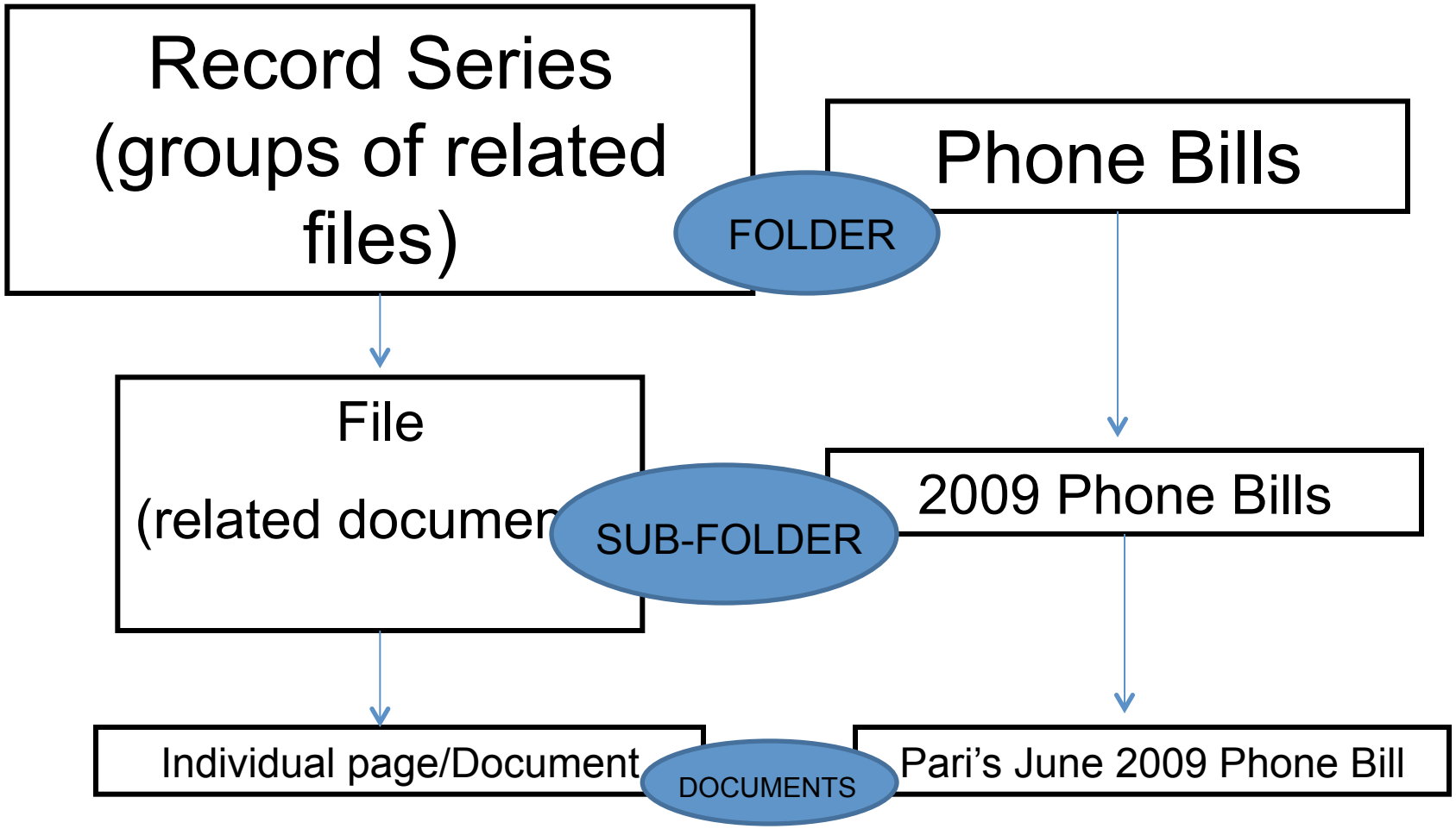


FUNCTIONAL FILING





Terminology: Record Series



Filing Email & E-records

It is less time consuming in the long-run to file email right away while the content is fresh.

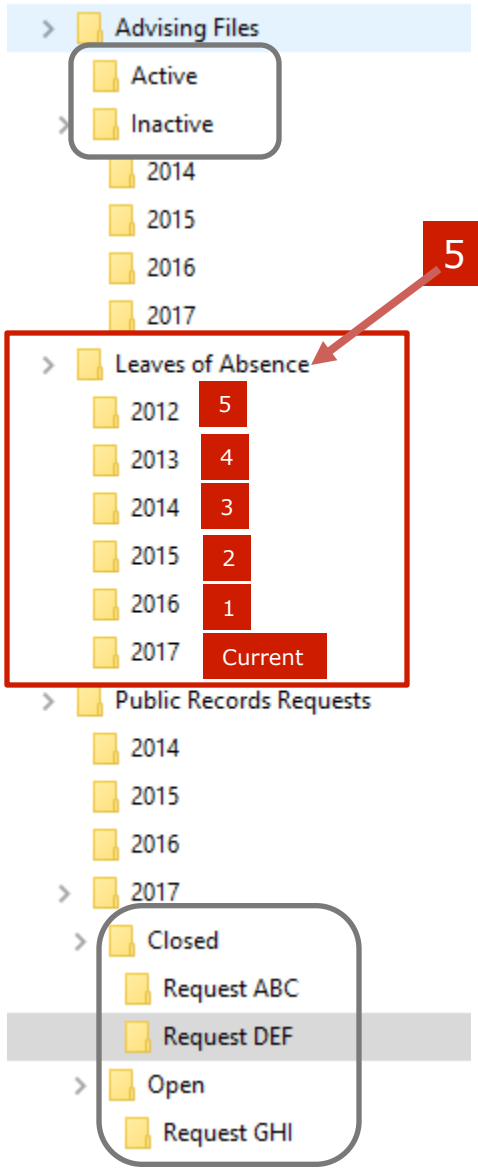
- Easier to create the basic file structure in advance, then drag and drop or save immediately
- Content is organized and accessible, even if you are not present
- Retention/disposition is easier to apply
- No clicking and rereading each file

Filing Email & E-records

Folder Structure

- On a central server space
- Create clear filing structures on server and/or inbox that everyone in section or functional area can follow
  - Think about filing for *retrieval* rather than creation
  - Folders are like index fields
- Incorporate records retention schedules into the folder names
  - Title of Folder = Title of Record Series (Per retention schedules)
- Subfolders can aid in organization, retention and disposition
  - When retention expires on a folder, right clicking once deletes all contents

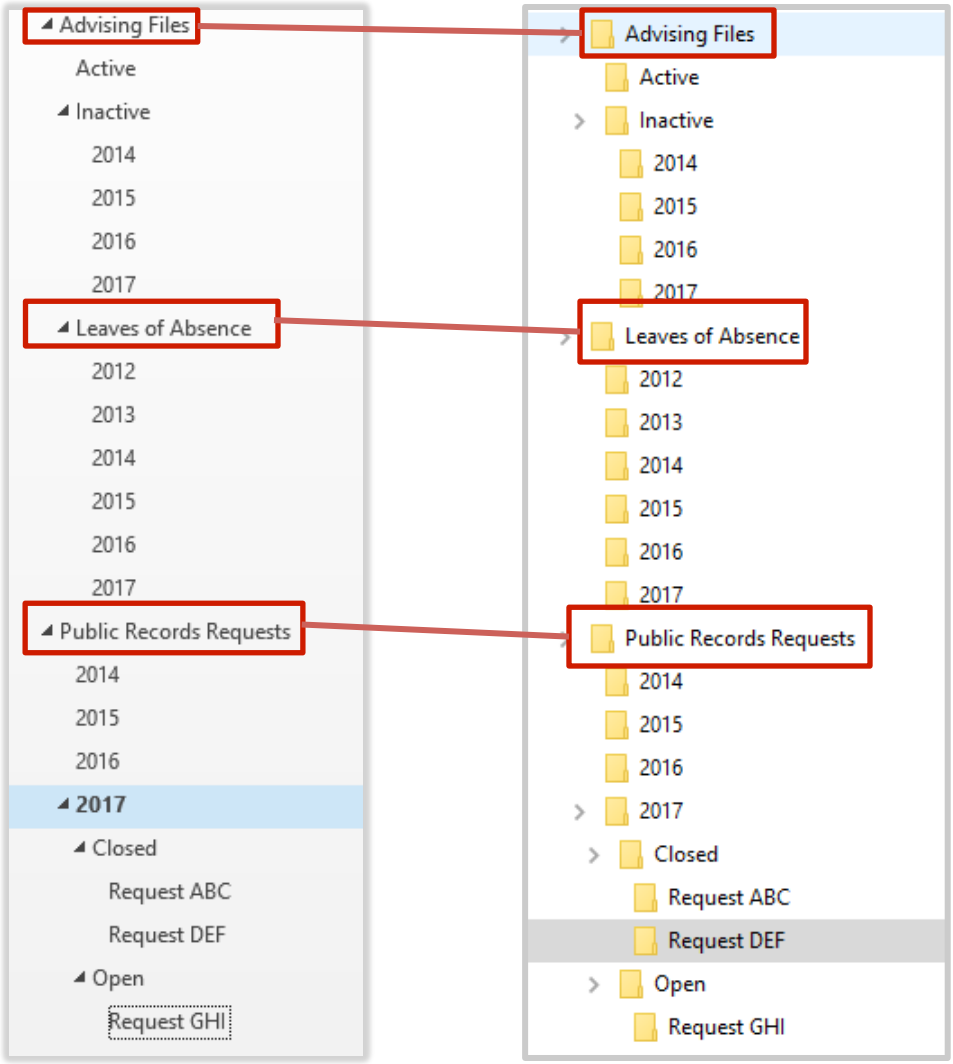
Structuring File Shares



Hint

You shouldn't have more annual folders than the retention period calls for, plus one for the current year.

Structuring Outlook



Structuring File Shares

NAMING DOCUMENTS

- Descriptive file names & naming conventions
  - Good: Memo_Office_Move or Jones_Case
  - Bad: Memo1 or AB_Case
  - Same structures to date file names
    - 2011_03_28 -or- 20110328
    - Put date in same location, before or after file name – this affects sorting
  - Consistent spellings, abbreviations, version style
    - Version2 –or- V2
  - Status Format
    - (superseded) –or- (final) –or- (draft)

Do Not use file formats as folder names.

PDF, Emails, Images

Consistency & Uniformity in naming and placement is key!

DELETING EMAIL AND ELECTRONIC RECORDS

DELETING EMAIL & E-RECORDS

It really is okay to delete…as long as it is done with retention schedules and consistently

- According to approved retention schedules
  - In conjunction with related hardcopy records (unless scheduled for different retention periods)

- Can be deleted immediately:
  - Transient Records retention schedule
  - For non-records

Why can't we just keep all email & e-records?

Increases

- Storage space
- Maintenance costs
- Retrieval time
- Public records requests
- Cost of e-discovery production and preservation
- Liability

Decreases

- Ability to quickly locate
- Defensibility
- Compliance

There is a line between business value and over preservation – retention schedules define that line upfront

When an Employee Leaves

- Email retains its “record” status
- If saved to functional file share it will be:
  - Searchable and retrievable
  - Deleted in accordance with proper retention schedule
- Manager should
  - Request access to employee’s mailbox
  - Request access to employee’s individual file space
  - Review as part of separation procedures

FACTORS IN MEDIA PERMANENCE

WHICH HAVE YOU EVER USED OR FOUND:

WHAT IS THE LONGEST RETENTION PERIOD FOR A RECORD SERIES IN THE OFFICE?

A. 5 years

B. 15 years

C. 50 years

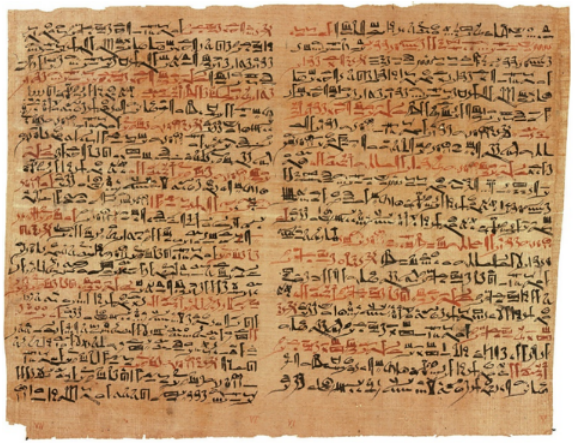
D. Permanent

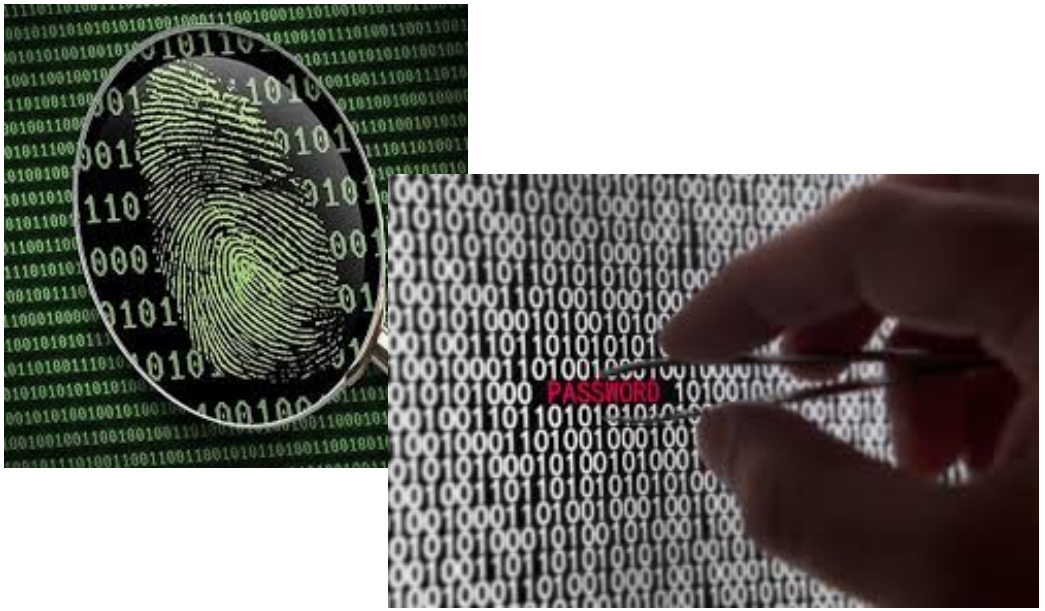
Why are some of your records permanent?

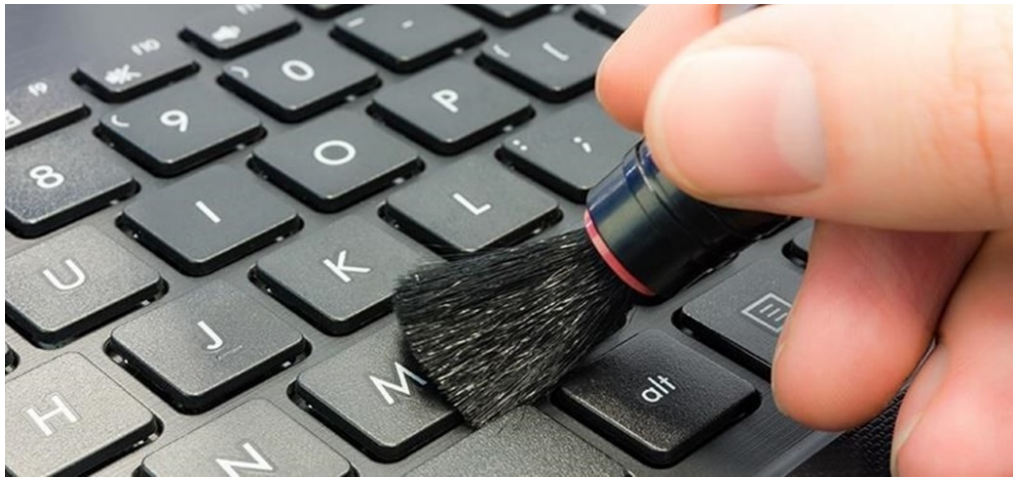
A. They're pretty to look at
B. We have the space, so why not?
C. Historically valuable
D. Legal record of actions taken
E. Our public records request stats need a boost

LIFE EXPECTANCY OF ANALOG FORMATS

1000s of years **if stored properly**




PASSWORD



DOMESDAY BOOK - 1086

Wooden chest, cased, lined and bound in iron and secured by three different locks, in which Domesday Book was kept stored from about 1600. Catalogue reference: E 31/4

DOMESDAY PROJECT – 1986 (BBC)

- 1 million people submitted information
- 150,000 pages of text
- 20,000 photos
- 100s of maps, statistics, and videos
- Used LaserDisc technology
- Specially designed interface to search using menus, keywords, maps and more.
- By 2002 (15 years), it was inaccessible
  - LaserDisc lost out to CDs, so they stopped manufacturing discs and readers
  - Software to read binary code was not preserved
- While some has been painstakingly reconstructed, the menus and windows have not, thus taking away from the intended user experience.

Google Cloud Storage - 2009

- Software upgrades in data centers are delicate
  - Performed during routine operations
  - Without downtime for customers
- 1000s of Gmail boxes deleted during software upgrade

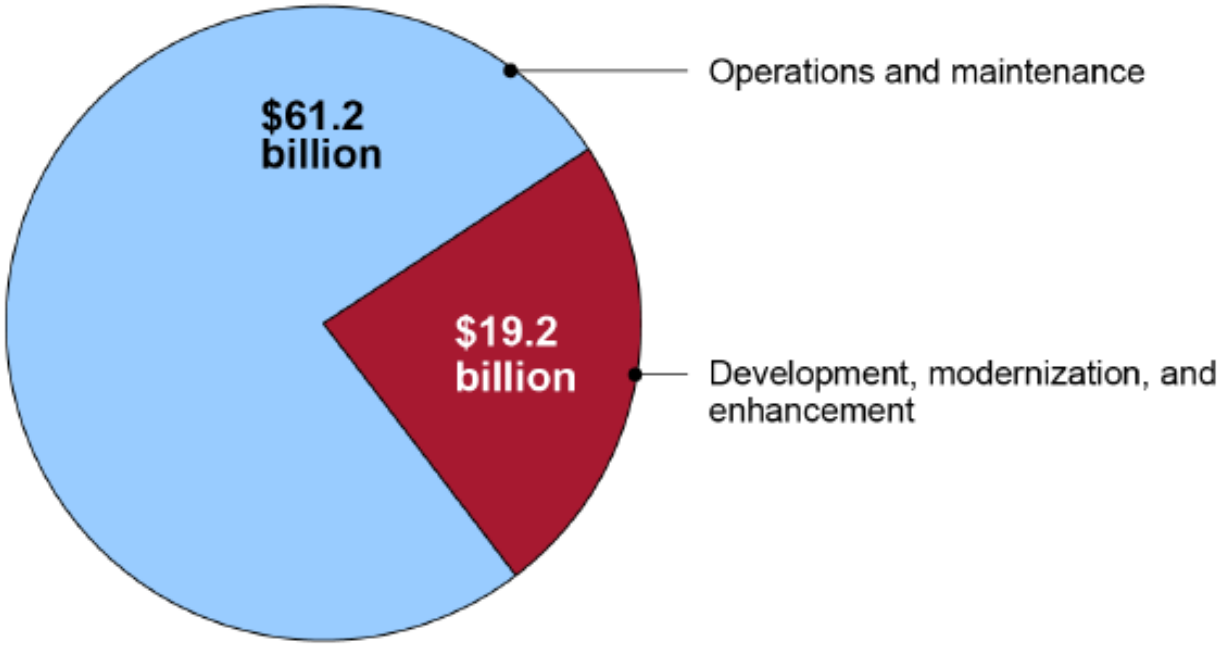
Amazon Data Center - 2011

- Technical glitch caused loss of 0.07% of the information they were storing
- Hundreds of thousands of gigabytes

- What if a government entity accidently lost that much information
  - Public records
  - Legal records
  - Historical records

US GOVERNMENT ACCOUNTABILITY OFFICE: *FEDERAL AGENCIES NEED TO ADDRESS AGING LEGACY SYSTEMS*

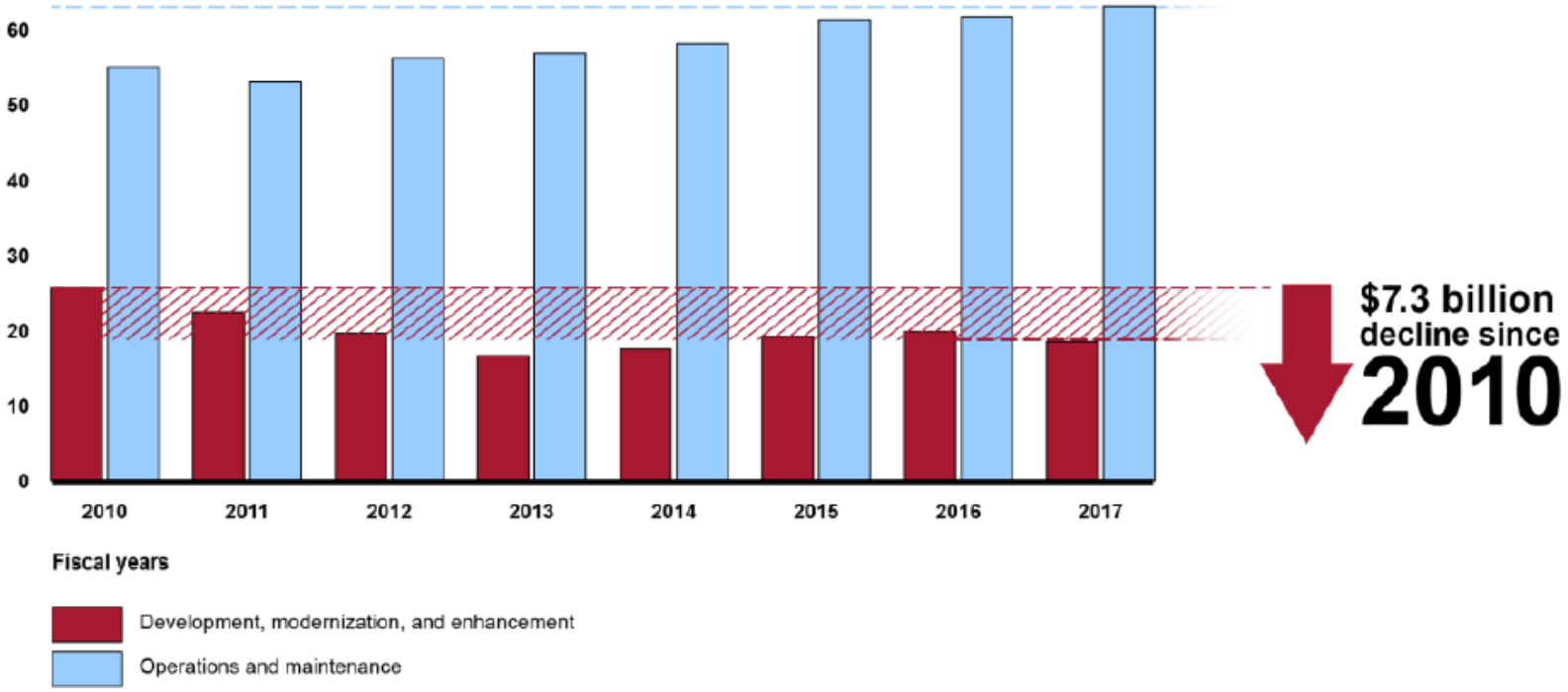
Figure 1: Fiscal Year 2015 Federal Spending on IT Operations and Maintenance and Development, Modernization, and Enhancement



Source: GAO analysis of Office of Management and Budget's Information Technology Dashboard | GAO-16-696T

US Government Accountability Office: *Federal Agencies Need to Address Aging Legacy Systems*

Figure 2: Summary of IT Spending by Fiscal Year from 2010 through 2017 (Dollars in Billions)



Source: GAO analysis of agency data. | GAO-16-696T

IT'S ALL ABOUT

PRESERVATION!

Ensuring Access: 3 Goals

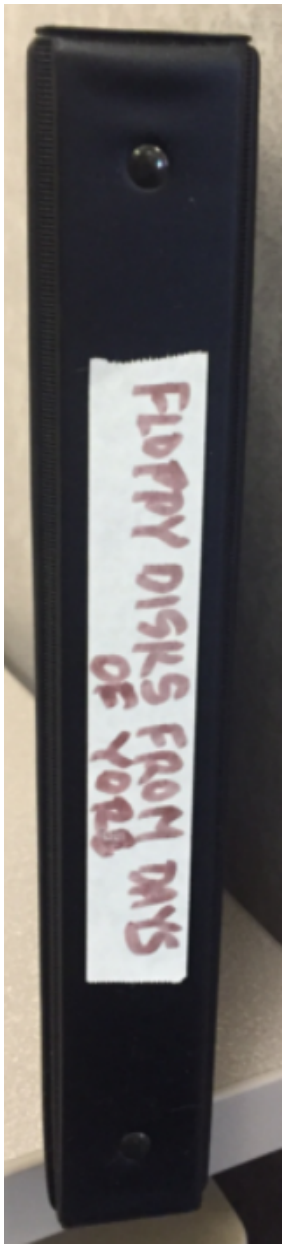
1. Availability

2. Readability

3. Reliability & Authenticity

GOAL 1: AVAILABILITY

The record is kept for its full retention period



Goal 2: Readability

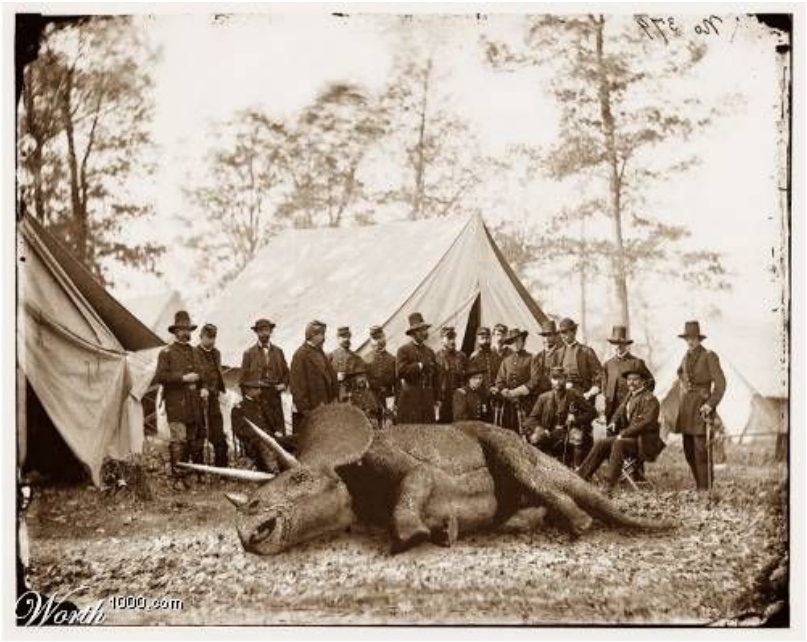
✓Available

xReadable

Goal 3: Reliability & Authenticity

- Record is what it claims to be
  - Reliability – full and accurate representation
- Record is complete and unaltered
  - All metadata
- Need audit trails
  - What was changed?
  - By whom?
  - When?
  - Documentation of migration(s)

AUTHENTICITY = TRUST


Worth1000.com

How is it that these clues survived 200 years?

How would the story have changed?

Here's Why

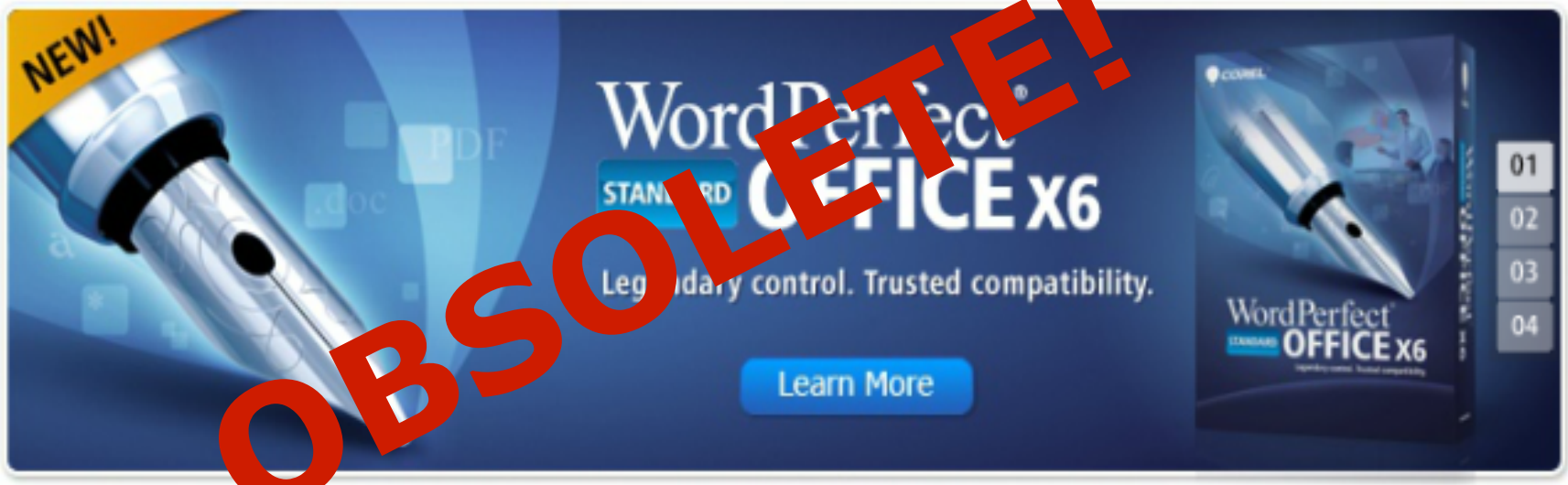
Probably had no idea that the flash drive:

- Stored records
- Contained something worth retrieving

But let's say Ben's history prowess told him the flash drive was once an information storage device…

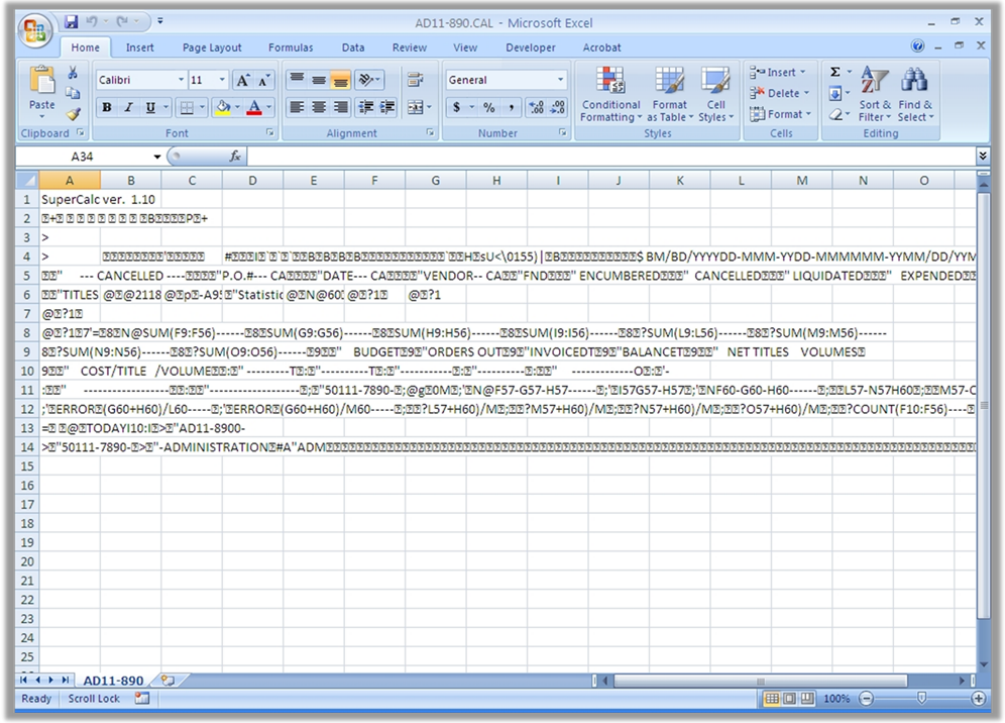
- Decay of physical media
- Hardware obsolescence

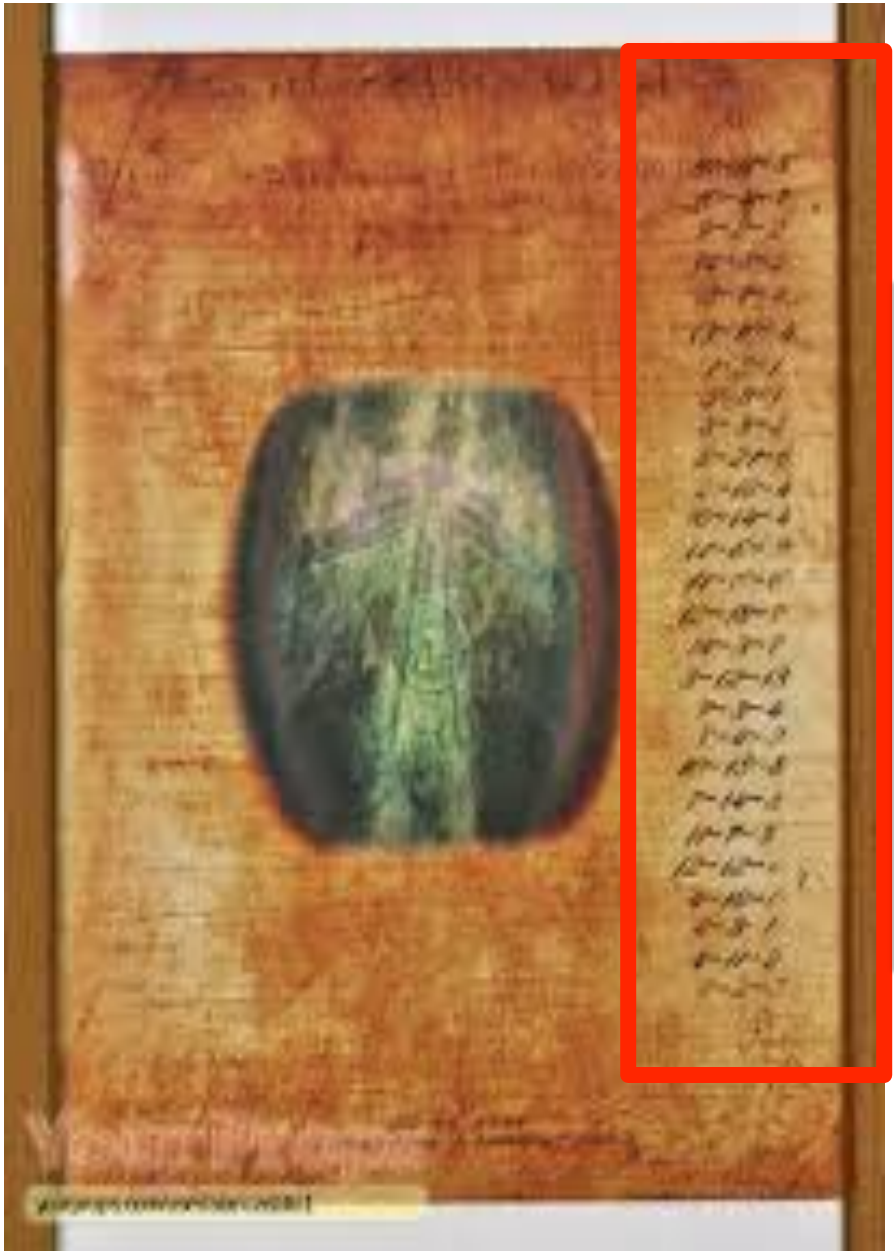
- Loss of information about format
- Software obsolescence

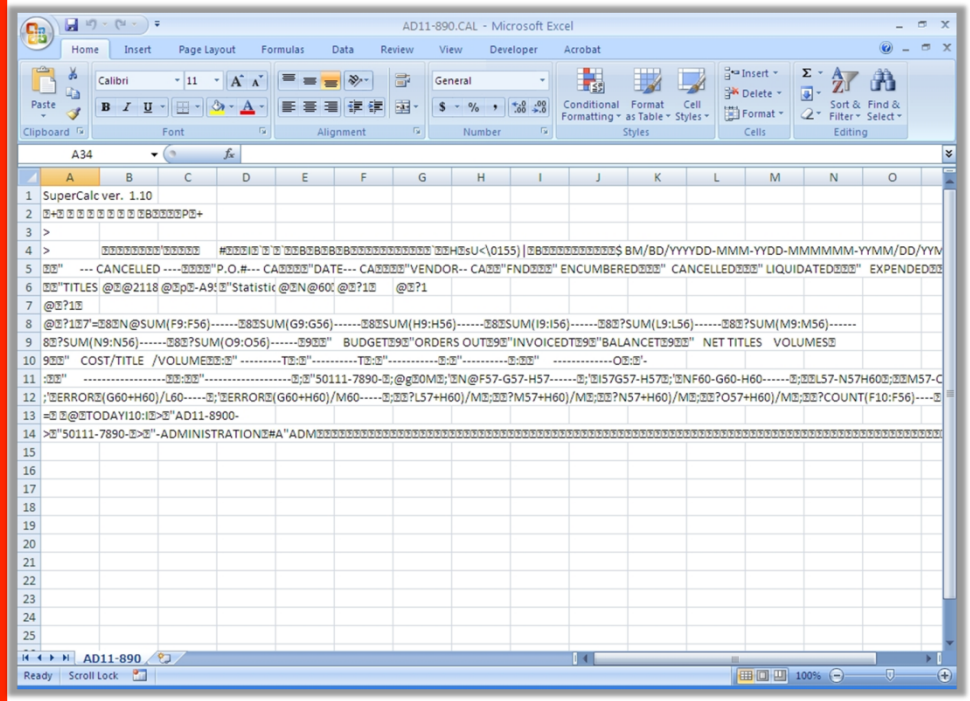


Content may still exist

But structure lost

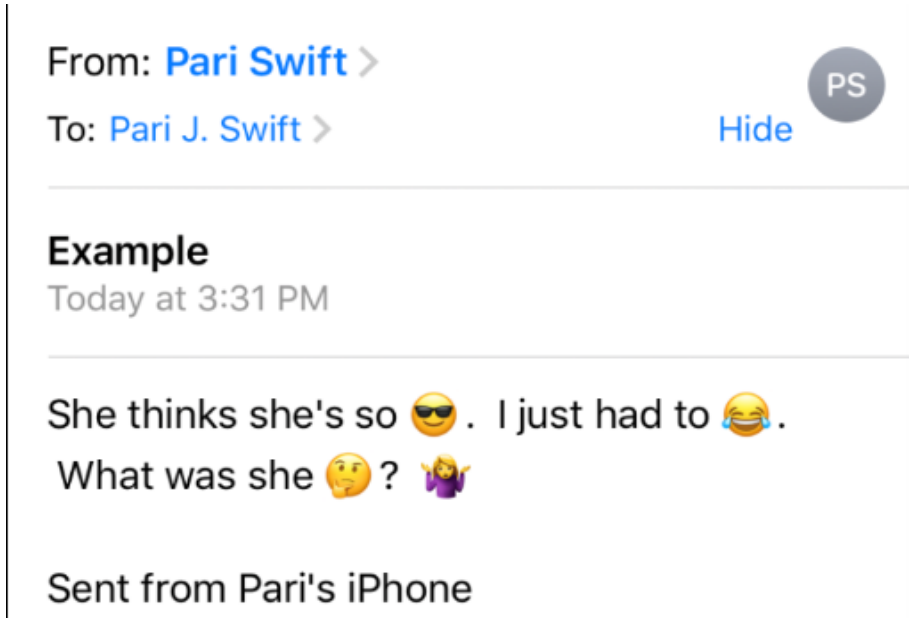




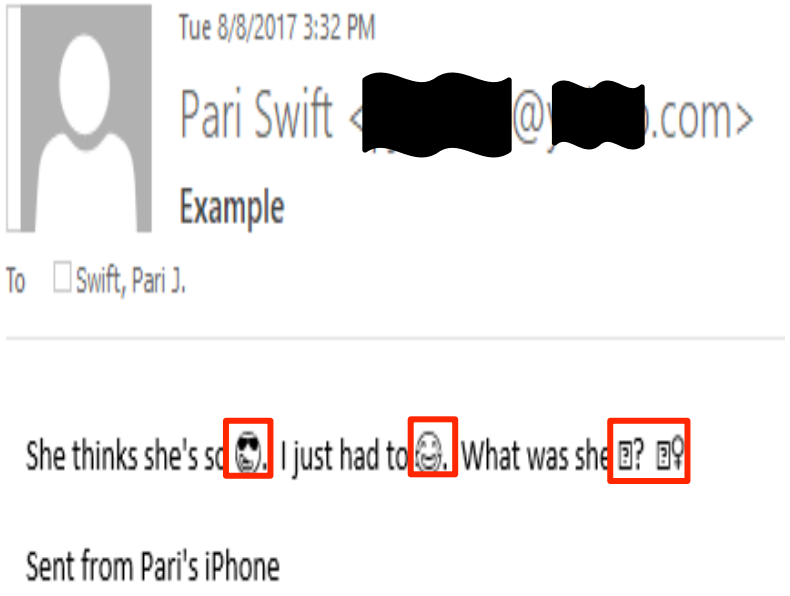

AD11-890.CAL - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
Acrobat
Paste
Calibri
11
General
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
Sort & Filter
Find & Select
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
A34
SuperCalc ver. 1.10
AD11-890
Ready
Scroll Lock
100%

Binary Code

Character	Binary Code	Character	Binary Code	Character	Binary Code
A	01000001	a	01100001	!	00100001
B	01000010	b	01100010	"	00100010
C	01000011	c	01100011	#	00100011
D	01000100	d	01100100	$	00100100
E	01000101	e	01100101	%	00100101
F	01000110	f	01100110	&	00100110
G	01000111	g	01100111	'	00100111
H	01001000	h	01101000	(	00101000
I	01001001	i	01101001	)	00101001
J	01001010	j	01101010	*	00101010
K	01001011	k	01101011	+	00101011
L	01001100	l	01101100	,	00101100
M	01001101	m	01101101	-	00101101
N	01001110	n	01101110	.	00101110
O	01001111	o	01101111	/	00101111
P	01010000	p	01110000	0	00110000
Q	01010001	q	01110001	1	00110001
R	01010010	r	01110010	2	00110010
S	01010011	s	01110011	3	00110011
T	01010100	t	01110100	4	00110100
U	01010101	u	01110101	5	00110101
V	01010110	v	01110110	6	00110110
W	01010111	w	01110111	7	00110111
X	01011000	x	01111000	8	00111000
Y	01011001	y	01111001	9	00111001
Z	01011010	z	01111010	?	00111111
				@	01000000



Screenshot of email sent through Yahoo.com from iPhone using iOS 10.3.3



Screenshot of email as received in Outlook 2016

From: Pari Swift >
To: Pari J. Swift > Hide

PS

Example
Today at 3:31 PM

She thinks she's so 😎. I just had to 😂.
What was she 🤔? 🤷‍♀️

Sent from Pari's iPhone

Email sent through Yahoo.com from iPhone using iOS 10.3.3

Pari Swift <pjsatohs@yahoo.com>
To Pari J. Swift

She thinks she's so 😎. I just had to 😭 What was she 🤔? 🤷‍♀️

Sent from Pari's iPhone

Email sent through Yahoo.com viewed in Chrome browser

```
Content-Type: text/plain;
        charset=utf-8
Content-Transfer-Encoding: quoted-printable
From: Pari Swift <pjsatohs@yahoo.com>
Mime-Version: 1.0 (1.0)
Date: Tue, 8 Aug 2017 15:31:33 -0400
Subject: Example
Message-Id: <0BF9A6BF-5B31-4EDA-AD7E-F908A1E739E3@yahoo.com>
To: "Pari J. Swift" <swift.102@osu.edu>
Content-Length: 163

She thinks she's so =F0=9F=98=8E  I just had to =F0=9F=98=82.  What was she=
 =F0=9F=A4=94? =F0=9F=A4=B7=E2=80=8D=E2=99=80=EF=B8=8F=20

Sent from Pari's iPhone=
```

Yahoo email *source code* viewed through:

- Internet Explorer
- Microsoft Edge
- Chrome

WHAT DOES THIS ALL MEAN?

- Technology will change.
- Media will degrade.
- It's **your** job to make sure that the record is still **available**, **readable**, and **authentic** – until it's met retention.

DIGITAL PRESERVATION IS THE CHALLENGE OF MAINTAINING THESE CHARACTERISTICS OVER TIME

SO, HOW DO I THINK THE STORY WOULD HAVE CHANGED?


Credit: National Trust for Historic Preservation


What do you care? You got the girl.
Abigail: It's true.
Ben: It's true.

PERMANENT ELECTRONIC RECORDS

Just because the law allows for it

doesn't mean it is

LIABILITY

- Unstable media
- Easily Alterable
- Degrade over time (might not be visible)

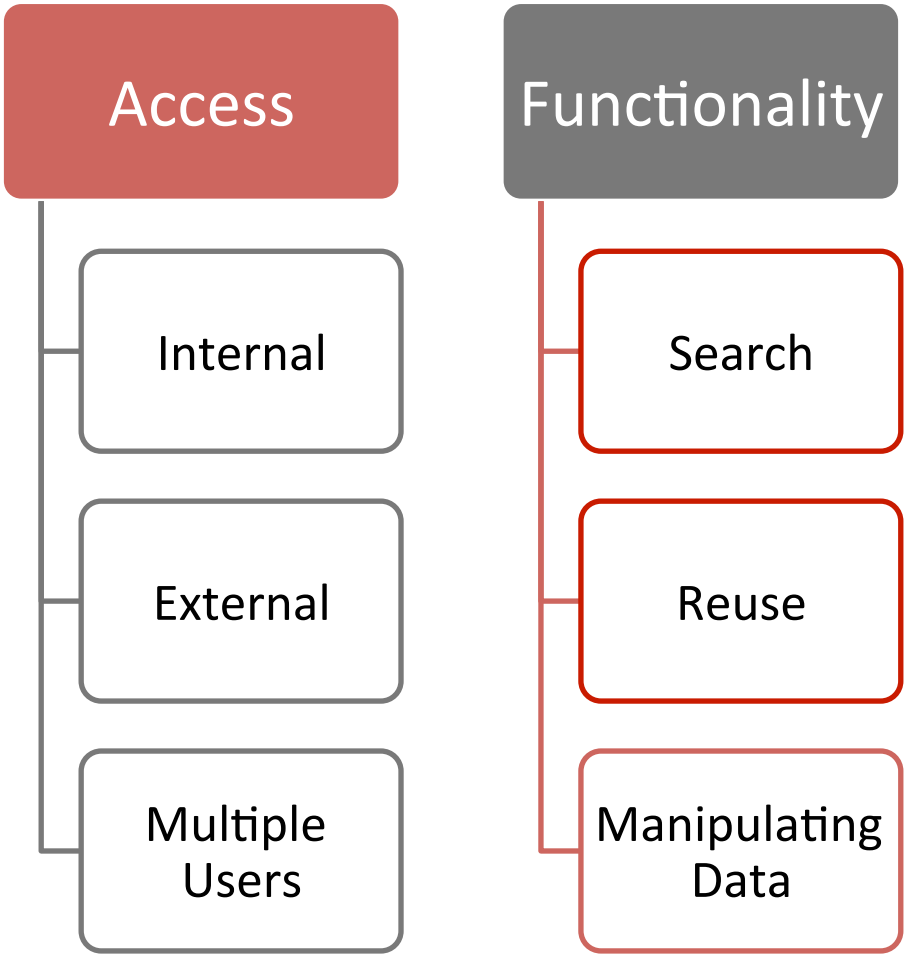
Yet…

If retention=permanent, your office has an obligation to always (forever) be able to:

- Provide access
- Prove reliability and authenticity
- Just having the data isn't enough!

And if you can't…

Electronic Storage is Great For:



TO SCAN OR NOT TO SCAN

DIGITIZATION FACTORS TO CONSIDER

OhioERC Scanning Feasibility Tool

5. What are the types of records?

6. How many records are you reviewing/accumulating for possible scanning?

a) 1-10 Boxes
b) 11-20 Boxes
c) 21-50 Boxes
d) 51 or More Boxes

Answer

Answer

www.ohioerc.org

- Guidelines and Tips
  - Document Imaging

OhioERC Scanning Feasibility Tool: Benefits

- Workflow or Process Changes
- Cost-Benefit Applications
- Manage storage and server costs
- Answer Summaries
- Final Report for making case

DIGITIZATION FACTORS TO CONSIDER

- Scan Format
- Scanning Resolution
- Verification
- Metadata
- Indexing
- Storage/ Backup
- Migration/ Media Refreshing
- Process Audits

Cost Justification

Invalid Reasons to Scan

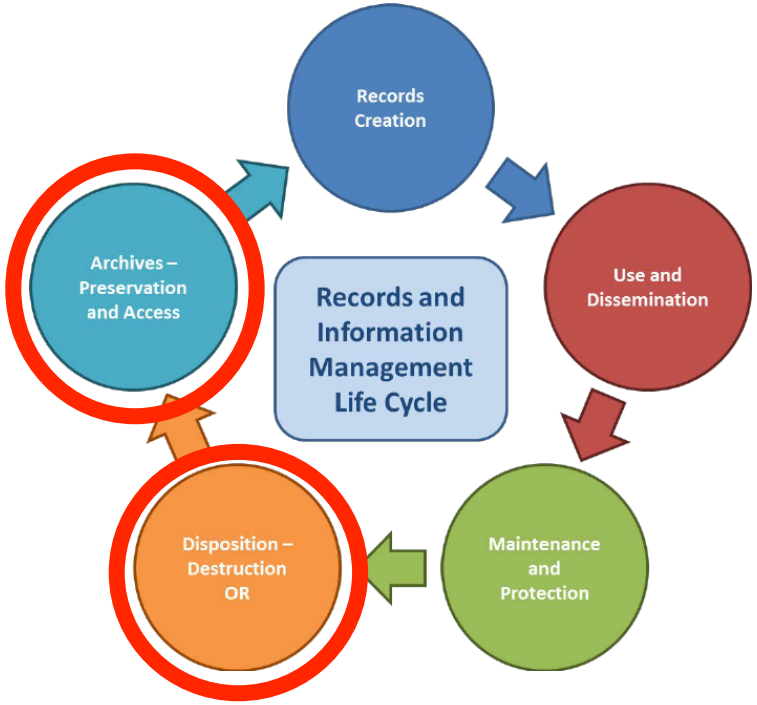
- Physical space problems
- Paperless for the sake of being paperless
- Low Reference documents
- Documents nearing end of retention

Valid Reasons to Scan

- Enhanced Business Process
- Multiple Points of Access
- Consolidated Recordkeeping
- Integration with Enterprise Systems
- Single Copy Management

INFORMATION MANAGEMENT LIFECYCLE

- Can it purge?
- Tools for each step
- What preservation standards do the vendors use?
- Missing a step in the technology process could mean losing access to information
- Analog back-up



Preservation Strategy: Convert to Hardcopy

- Viable solution when:
  - All necessary metadata can be captured
  - No reason for electronic functionalities
  - Use goes down over time
  - Color not as important

- Ideal for text-based records
  - Text files, word processing
  - Imaged paper
  - Photos

Digital Preservation

WHEN RECORDS ARE AT RISK

Software is obsolete or nearing end of life

If it is dependent upon specific hardware or systems

Proprietary or specialty formats

When systems are being replaced

When the records are no longer actively used

Any records on removable media

ARCHIVING VS. ARCHIVING VS. BACKUP

Historical

- Accumulation and preservation of *historical* records
- *Permanent* value
  - Cultural
  - Historical
  - Evidentiary

IT Archive

- File that no longer changes
- Frees up disk space
- Slower search speed
- Searchability critical
- Retained for compliance
- Might have longer-term retention
- *Original* – removed from initial location and stored elsewhere
- Added metadata
- Specialized software has integrated conversion capabilities

IT Backup

- Primarily used for operational recoveries
- Speed is important
- *Copy*
- Deleted after replaced

An IT Backup is not a sufficient digital preservation strategy.

Electronic Recordkeeping Systems

- What they do:
  - Organize and Categorize
  - Centralize control
  - Improve access
  - Storage

- They are not preservation systems. They don't:
  - Provide for format validation
  - Migration

Digital Preservation Strategy

- Ensures long-term
  - Accessibility
  - Authenticity
  - Integrity
  - Trustworthiness
- For the Purpose of
  - Operational mandates
  - Audits
  - Evidence
  - Research

Digital Preservation seeks to manage records so that they will remain usable through many successive generations of technological advancement.

Digital Preservation Strategy

- Long-term records are:
  - Tracked
  - Validated
  - Protected
  - Migrated

Resources:
- CoSA PERTTS Portal https://www.statearchivists.org/pertts/
- Library of Congress Digital Preservation http://www.digitalpreservation.gov
- MIT Libraries Digital Preservation Management http://www.dpworkshop.org

- Addresses issues of:
  - Software and hardware obsolescence
  - Security and file integrity
  - Needs of different user groups over time

Digital Preservation Strategy

There are no permanent electronic storage media.

- Proac oid catast
  - Med
  - Migr ap is too wide
  - Eve uman attention
  - No a
- Best ti vation strategy is at the time of creation.

WHAT IN THE WORDLE IS THIS?

ON-GOING COSTS OF ELECTRONIC RECORDS MANAGEMENT

Salaries & benefits

- IT
- Records Management
- End users

Hardware maintenance, upgrades, replacement

- Servers
- Storage devices
- Computers and accessories
- IT infrastructure (networks, etc.)

Software licenses, maintenance, upgrades

- Server operating system and database
- PC operating system and desktop
- Document/image/records management
- Security
- Backup

Training

Consultants

Migration and storage

https://www.archives.gov/records-mgmt/policy/costs-and-benefits-erm-erk.html

ONE-TIME COSTS OF ELECTRONIC RECORDS MANAGEMENT

Salaries & benefits

- Software conversion
- Data conversion
- Analysist, design, development
- Testing

Hardware – new equipment or infrastructure

- Servers
- Storage devices
- Computers and accessories
- IT infrastructure (networks, etc.)

Software – new

- Server operating system and database
- PC operating system and desktop
- Document/image/records management
- Security
- Backup
- Customization

Training

- End user
- System administrator
- Change management

Consultants

https://www.archives.gov/records-mgmt/policy/costs-and-benefits-erm-erk.html

Long-Term Preservation Strategy: Standard Digital Formats

- Formats expected not to change (or change slowly)
- Widely supported and used
  - (otherwise the format will cease to exist)
- Easily Transferable
  - Compatible with other applications
  - Forward/Backward compatable
  - Version control
- Non-proprietary/Open/Published standard
  - (other companies can use the standard)

Standards
- ANSI
- ISO
- Industry

"…digital information lasts forever—or five years, whichever comes first."

Jeff Rothenberg
Ensuring the Longevity of Digital Information

WHO SHOULD BE AT THE TABLE?

CONCLUSION

- We can't afford to say we'll leave our history to the next office-holder.
- We need to act at the beginning of the information lifecycle to ensure that historical information survives.
- We need to:
  - Inform vendors of our concerns
  - Have the right people in the team during selection
    - Someone representing archives and preservation
    - Someone representing records management
  - And ask questions about the long-term preservation of records created by, stored on, or read by their products
  - Use the guidance that is available

Existing OhioERC Guidance

- Electronic Records Management
- Legal Obligations
- RIM Role in E-Discovery
- Cloud Computing
- Databases as Public Record
- Digital Document Imaging
- Managing Email
- Hybrid Microfilming
- Social Media
- **Scanning Feasibility**
- Tip Sheets

www.ohioerc.org

TYPES OF PRESERVATION TOOLS

- What file formats to you have?
  - DROID
    - Run against directories to determine what file formats you have
    - Identifies everything that you have by format
  - File Information Tool Set (FITS)
    - Identifies formats
    - Validates files
    - Extracts technical and embedded metadata
- Open Formats
  - Open Office
    - LibreOffice
    - Apache Open Office
  - Text Reader
    - NotePad++
  - VLC Media Player
- Data Integrity
  - Checksums
    - Creates a digital footprint
    - When run subsequent times, tells you if there are changes
      - But not what was changed
      - And not if the change was bit-rot or intentional
    - Tools
      - Bagger
      - ExactFile
      - Fixity
- Preservation repository software
  - Achivematica/archivesDIRECT
  - Dspace
  - Islandora/Fedora
  - Preservica
  - DSPS
  - AXAEM
  - To name a few

Digital Preservation Guidance and Tools

- National Archives and Records Administration
  - *NEW Universal Electronic Records Management Requirements
  - Format Recommendations
  - Electronic Records Archives
  - Electronic Records Management Automation
  - Federal Requirements for Including Recordkeeping in Agency Electronic Information Systems
  - Toolkit for Managing Electronic Records
  - Records Management Language for Contracts
- Library of Congress
  - Sustainability Factors for Digital Formats
  - Sustainability of Digital Formats
- National Digital Stewardship Alliance
  - Levels of Digital Preservation
- National Archives of the UK
  - DROID File Profiling Tool
  - PRONOM Technical Registry
- Council of State Archivists
  - PERTTS Portal (Program for Electronic Records Training, Tools and Standards)
  - State Electronic Records Initiative (SERI)
- ARMA International
  - Updated Guide to Commonly Used U.S. Nations and International Records Management Standards and Best Practices. (www.armaedfoundation.org)
- State Guidelines
  - North Carolina
  - Texas
  - Kentucky
  - Missouri

PARI J. SWIFT
UNIVERSITY RECORDS MANAGER
THE OHIO STATE UNIVERSITY
SWIFT.102@OSU.EDU
614-292-4092