

Preserving Electronic Records in Colleges and Universities

Getting Your Program Off the Ground

Presented by the NYS Archives

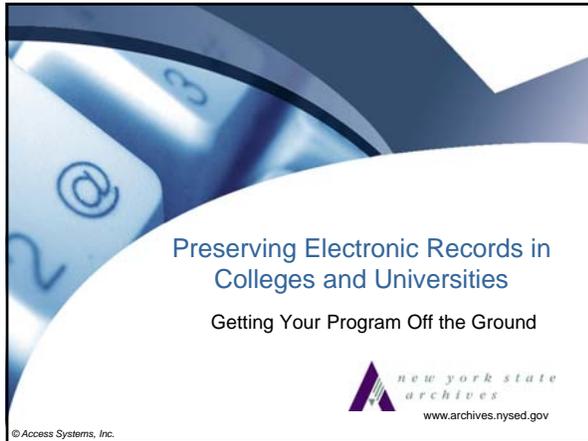
Agenda and Presentation Slides



Preserving Electronic Records in Colleges and Universities

AGENDA

<i>Section One</i>	Welcome & Introductions
<i>Section Two</i>	Building the foundation
<i>Section Three</i>	E-Records in your environment & discussion
<i>Section Four</i>	Awareness of the issues & discussion
<i>Section Five</i>	Preservation examples, standards & discussion
<i>Section Six</i>	Setting goals and suggested strategies
<i>Section Seven</i>	Disaster preparedness for e-records
<i>Section Eight</i>	Developing an action plan
<i>Section Nine</i>	Summary and wrap up



Preserving Electronic Records in Colleges and Universities

Getting Your Program Off the Ground



new york state archives
www.archives.nysed.gov

© Access Systems, Inc.

What is offered with this workshop

- An understanding of the issues
- Strategies to follow
- High level action plan
- Additional resources to help you

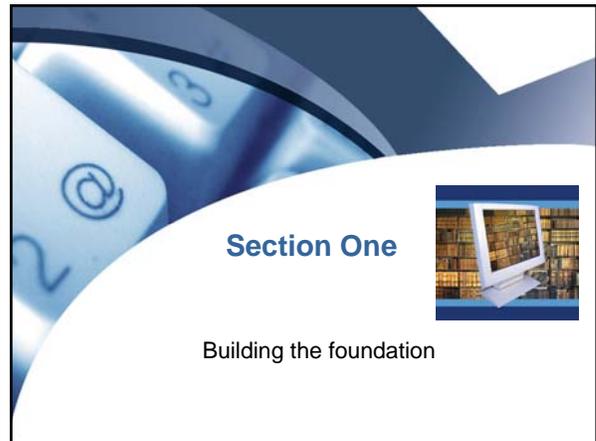
2

Workshop outline



1. Building the foundation
2. E-Records in your environment
3. Awareness of the issues & file formats
4. **Standards & e-preservation initiatives**
5. Goals and strategies
6. Disaster preparedness
7. Action plan
8. Summary & wrap-up

3



Section One



Building the foundation

Whether on paper or in electronic bits, the future readability of the media is the concern



5

Goal of preservation

- Ensuring the use and accessibility of information in a record within the proper context for that record's full retention period



6

Digital preservation challenges

- Technological obsolescence
 - Hardware, software, formats, media vulnerability
- Organizational issues
 - Content 'ownership' issues
 - Competing stakeholder interests
- Legal/compliance issues
 - Authenticity, security
- Resource requirements
 - Staffing, equipment, space, capital

7

Growing awareness of preservation needs

- Domesday Book
 - Book commissioned by William the Conqueror in 1086
 - An inventory of eleventh-century England compiled by Norman monks

Wooden chest, cased, lined and bound in iron and secured by three different locks, in which Domesday Book was kept stored from about 1600. Photo from The National Archives of the UK.



8

Growing awareness of preservation needs

- Domesday project
 - In 1986, the BBC created multi-media presentation on a 12" Laserdisc
 - By 2002 the discs were unreadable
 - Hardware readers not available
 - Software not compatible with other devices
 - Concerns with media lifespan



9

Growing awareness of preservation needs

- Domesday result
 - Domesday Preservation Group
 - A group whose intent was to bring together people with an interest in reviving the Domesday project
 - No activity since one founder unexpectedly died in 2008
 - CAMILEON Project
 - Partnership between the University of Leeds and University of Michigan developed a system capable of accessing much of the discs using emulation techniques

10

And oh, by the way...

- Original book volumes are in fine condition in the National Archives in Kew, outside of London



11

Other examples or horror stories

- 1960 United States census data
 - Raw data was stored on magnetic tapes
 - By the late 1970s, the data were unreadable
- NASA satellite photography files 1970's
- College's decision to maintain old Student Management System backfires
 - Not all information transferred properly
 - Kept old system for lookup purposes
 - New IT Director decommissioned old system



Every 5 years, media becomes obsolete¹²

12

Let's hear from you...

- Share some of your own experiences
 - Not being able to read older records
 - Conversions gone bad, media failure, system malfunctions, lack of planning, etc.
 - Let's not leave out positive experiences too!



13

Section Two



E-records to be aware of in your environment

So what e-records should you focus on?

- Mandated to retain for compliance reasons
 - Permanent or require long-term retention
 - General guideline is to focus on those e-records with retention periods of greater than 5-10 years
 - Legally required by governing body
 - US Dept of Education, IRS, or other federal, state or local agency
 - Retention schedules adopted by your institution
 - NYS Archives, NARA, American Association of Collegiate Registrars and Admissions Officers (AACRAO), etc.

15

So what e-records should you focus on?

- Those records that involve litigation
 - Some matters can last many years or decades forcing retention of e-records and the system that created them
 - May wish to retain other primary documentation involving famous parties or significant matters



16

So what e-records should you focus on?

- Those with historical value to the institution

Plus other stuff you're 'told' to keep!



17

What are examples of permanent records?



18

Examples of permanent records

- President's Office
 - Files with significant subject or major policy-making program development process
- Student records
 - Transcripts
 - FERPA - Directory Information Policy Statement
 - List of scholarships awarded students

19

Examples of permanent records

- Academic affairs
 - Accreditation records (reports, questionnaires, self-study records, guides and related documents transmitted between accrediting bodies with significant importance)
 - Successfully registered academic program proposals
 - Curriculum registration records with State Education Department
 - Course information records (Official copy of student handbook, college catalogs, etc.)
 - Published faculty and student research records

20

Examples of permanent records

- Alumni
 - Alumni directory
- Other administrative records
 - Chartering documents and establishment plans
 - Special occasion/event file – Official copy of any program or promotional literature, photograph of events or games
 - Commencement records – official program or publication
 - Annual security report
 - Press releases

21

What could be 'longer term' records?

- 'Event' based retention periods
 - Student records for the 'perpetual' student
 - Retention could be 6 years after graduation or date of last attendance, but what happens when a student takes at least one course every 6 years?
 - Financial aid data in Student Information System
 - Could be up to 6 years *after* the loan is repaid
 - Student activity or organization records of historical significant
 - Contracts (could be 6 years from *expiration*)

22

Other types to be aware of (with potential long term retention?)

- E-mail
 - Created by senior administrators (e.g. president, deans, department heads, etc.)
 - Dependent on contents of e-mail
- Multimedia
 - Audio recordings, web cam events, video recording of significant lectures & presentations
- Web site(s)
 - Public facing (Internet) and internal (intranet)

23

More examples in the digital landscape

- Published articles, e-books, technical reports, journals
 - e.g. PDF, Word, scanned documents
- Electronic lab notebooks
 - e.g. Scanned images, proprietary database structure, and/or multiple individual files for one notebook (e.g. Word document with Excel charts which support modeling data sets from proprietary programs)
- Theses and dissertations



24

More examples in the digital landscape

- Conference papers & presentations
 - e.g. PDF, Word, PowerPoint, audio and video formats, Flash formats
- Learning objects from Learning Mgt System (LMS)
 - e.g. past courses and presentations stored in online learning system (e.g. Blackboard)
 - Could include core course, plus discussion threads, reference material and other files in various formats
 - SCORM (Sharable Content Object Reference Model) compliant systems are designed to transfer content from one compliant LMS to another

25

More examples in the digital landscape

- Source code from internally developed programs
 - e.g. legacy financial systems, databases and other application program's code
- Visualizations, simulations, & modeling datasets
 - e.g. proprietary applications conducted in research environments
- Social media records
 - Institution sanctioned wikis, Facebook pages, etc.



26

What other records do you see?

(or what records *did you see*, but not anymore?)

27

Section Three

Awareness of the issues



Traditional preservation areas of concern

- Accessibility
- Authenticity
- Provenance
- Security
- Operations
- Legal/compliance

Same issues for electronic files



29

...but e-records create additional challenges

30

More people, more devices creating records

As easy as this:



Using any of these:



31

More copies are being created



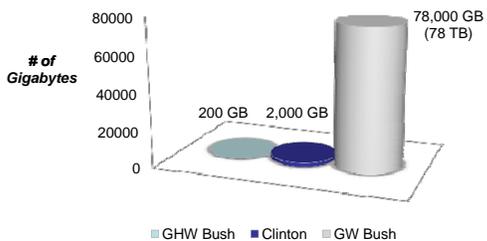
...and stored in many more places

and some copies have legs ...

32

Leading to tremendous growth in e-records

Approximately 85% of all government documents are created electronically



Administration	Storage Capacity
GHW Bush	200 GB
Clinton	2,000 GB
GW Bush	78,000 GB (78 TB)

Source: NARA

33

But growth is not always so obvious

Out of sight, out of mind

Can you tell which are the important records?



34

Leading to 'awareness' challenges

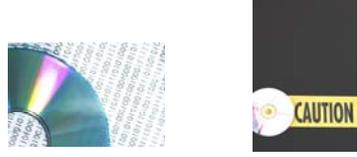
1. Awareness of what's out there
2. Awareness of the preservation issues
 - Technology
 - Operations (policy, procedures, legal, workflow, etc.)
3. Making others aware of all the issues
 - From senior officials to end users



35

Technology affects on e-readability

- Storage medium viability
 - Durability (e.g. 5 1/4" floppy disk vs. DVD-R)
 - Obsolescence (e.g. punch card, Laserdisc, etc.)



36

Name the year

	<u>Product</u>	<u>Introduced</u>	<u>Obsolete</u>
	Punch card	1930	?
	8" disk	1970	?
	5 1/4" disk	1975	?
	3.5" disk	1982	?
	CD	1988	?
	DVD-R, DVD-RW	1998	?
	USB Flash Drive	2001	?
	What's next?	?	??





37

What else affects e-readability?

- No device to read the medium
 - Need proper hardware
 - e.g. computer, disk drive, etc.
 - Proper interface to connect reading device to computer
 - USB port, FireWire, etc.




38

What else affects e-readability?

- Software version
 - Operating system (e.g., Windows, Mac, Linux, etc.)
 - Software application (e.g., Microsoft Word, Excel, Adobe Acrobat, etc.)
 - Software driver to interface with storage device
 - Versions, updates, and patches to all the above



39

What else affects e-readability?

- File format
 - Does your version of the software read the file format the information is stored within? (any Word 2.0 or 3.0 files?)



40

What file format is best?

- What format *guarantees* readability over the life span of a record?

None, but there are ways to lower the risk

41

Primary file format concerns

- Using a proprietary file format
 - Single company owns/controls the format
 - Will the company be around to support the format?
 - Will the company support older versions as technology changes?
- Using a format that compresses the file by removing data

42

Why use compression?

- Reduces size of original file
 - Scanned B&W image from >1MB to 35KB
 - Color image from >8MB to 2MB-80KB (depending on approach)
- Approaches
 - Lossless = no data lost when compressed
 - Lossy = some data lost when compressed
 - As a result, lossy is not recommended



43

Examples of different file formats

Format	Type	Compression	Comments
DOC	Word (Microsoft)	N/A	Proprietary but widely used
PDF	Text/Image	Lossy & Lossless*	Standard
PDF/A	Text/Image	Lossless*	Standard
TIFF	Image	Lossless	Standard
JPG	Image	Lossy	Standard
MPEG	Video	Lossy & Lossless	Standard
WMA	Video (Microsoft)	Lossy & Lossless	Proprietary but widely used
WAV	Audio	Uncompressed	Standard

* Depends on method used and content (e.g. is image included in PDF?)

44

Sample of the many different flavors of PDF

- PDF image
 - scanned
- PDF text
 - e.g. created from MS word
- PDF text & image
 - Scanned and OCR'd
- PDF/A
 - Open archival



45

PDF/A

- Subset of PDF for long-term preservation
 - Based on PDF 1.4
- Non-proprietary standard (ISO 19005)
 - ISO = International Organization for Standardization
- Not dependant on a particular viewer/reader
- Self-contained
 - Information needed for display is embedded
 - e.g. all content (text, images), fonts, color information

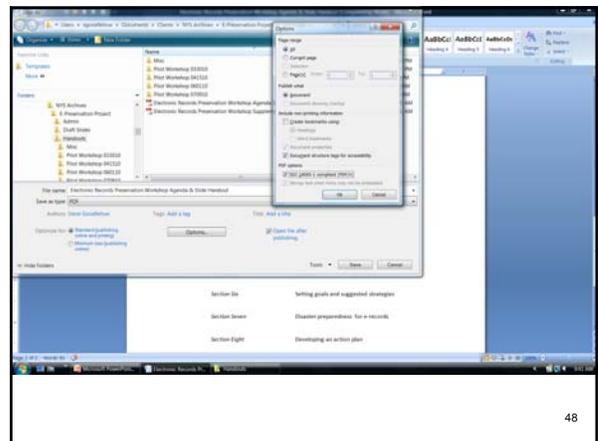
46

PDF/A

- What's not allowed in a PDF/A file:
 - Audio & video content
 - Information from external sources
 - e.g. font programs, hyperlinks, executable file launches
 - Encryption
 - LZW compression
 - Embedded files



47



48

Databases

- Examples:
 - Microsoft Access, Microsoft SQL Server, Oracle, MySQL, IBM DB2, etc.
 - Note: some of these can be the 'backend' of your student management system, HR, or financial system (e.g. PeopleSoft)
- No adequate standards
- Issue with long-term functionality
- XML is an option



49

Extensible Markup Language (XML)

- XML stands for EXtensible Markup Language
- XML is a markup language much like HTML
 - Designed to *carry* data, not to display data
 - HTML focuses on displaying content & appearance
 - XML focuses on describing data and structure
 - Developed to transport data between different operating systems



50

Extensible Markup Language (XML)

- Requires another program to process and read data
 - Uses tags to indicate the structure of data
 - Tags are not predefined; must define your own
 - XML is designed to be self-descriptive
- Open format
 - Recommended by W3C



51

Advantages of XML

Strengths

- De facto Internet standard
- Widely used
- Navigable
- **Transferable**
- Nominal file size increase
- Revisable
- Content separated from rendering



Weaknesses

- No inherent protection from ability to revise, affecting integrity
- **Content separated from rendering**



Charles Dollar, University of British Columbia

52

So how important is appearance to you?

- An e-mail printed in Times Roman vs. Arial?
- How about a report with tables or an otherwise complicated layout?
- What about animation within a presentation?

53

Desirable file format characteristics

- **Device independent**
 - Can be reliably and consistently rendered without regard to the hardware/software platform
- **Self-contained**
 - Contains all resources necessary for rendering
- **Transparent**
 - Allows direct analysis with basic tools
- **Self-documenting**
 - Contains its own description

Charles Dollar, University of British Columbia

54

Desirable file format characteristics

- No technical protection mechanisms
 - e.g. encryption, passwords, etc.
- Publicly available specification
- Adoption and widespread use
 - May be the best deterrent against preservation risk

Charles Dollar, University of British Columbia 55

File format summary

- File format selection should be driven by recordkeeping requirements
- Avoid proprietary format from single vendor
- **Must require ability to be transferred**
- TIFF, XML and PDF/A are good choices



Charles Dollar, University of British Columbia 56

Other e-preservation challenges

- Internal awareness
 - Do you know what is out there?
 - Do you know what new technologies are being planned or implemented?
- What else have you found?
 - New systems being introduced?
 - Proprietary formats being used?



57

Section Four



Standards & other e-preservation initiatives

Who is doing what?

Do standards exist?



What research is going on?



Keep in mind this is still all relatively new stuff!

59

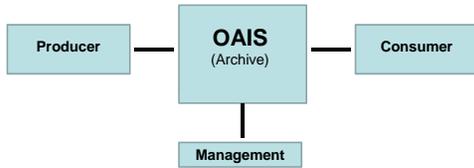
General framework

- Open Archival Information System (OAIS)
 - ISO conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term (www.oclc.org)
 - Purpose
 - Increase awareness and understanding of concepts
 - Create a framework to guide the identification and development of standards
 - Provide a common language for everyone



60

OAIS environment



Additional information on OAIS is available in the handout

61

Another example of standards

- Dublin Core Metadata Initiative
 - An open organization engaged in the development of interoperable metadata standards supporting a broad range of purposes and business models
 - <http://www.dublincore.org/>

Additional information on standards is available in the handout

62

Who is doing what?

• Research project examples

- LOCKSS (Lots of Copies Keep Stuff Safe)
 - Stanford University Libraries
- TIPR (Towards Interoperable Preservation Repositories)
 - Cornell (Fedora), NYU (DSpace), Florida Center for Library Automation (DAITSS)
- Chronopolis Digital Preservation Demonstration Project
 - San Diego Super Computer Center, USD Libraries, National Center for Atmospheric Research, University of Maryland Institute for Advance Computer Studies
- Rockefeller Archives Center/Smithsonian Institution Archives
 - Collaborative Electronic Records Project
 - Continued work under EMCAP (E-mail Collection And Preservation)



63

Sample of operational programs

- NARA's Electronic Records Archive (ERA)
- National Archives of Australia
- National Archives of the Netherland's Digital Longevity Testbed
- U of Florida Dark Archive in the Sunshine State (DAITSS)
- Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP)
- MetaArchive Cooperative
 - Uses LOCKSS



64

Sample of operational programs

- State governments
 - Arizona State Library, Archives & Public Records
 - Georgia Digital Archives
 - Kansas State Historical Society
 - NYS State Archives Electronic Records Unit
 - North Carolina State Archives
 - Minnesota Historical Society
 - Wisconsin Digital Archives
 - Washington Digital Archives



65

NYS Archives' Electronic Records Unit

- Relatively new initiative
- Follows preservation model developed by the National Archives of Australia
 - Quarantine station
 - Preservation station
 - Storage station



66

NYS Archives' Electronic Records Unit

- Records received match sent?
- Anti-virus check
- Checksum error detection

• Future conversion to normalized formats such as XML

- Store both original bit stream & converted versions
- Generate off-site backup copies

67

NYS Archives' Electronic Records Unit

- Continuing to research best practices
- Creating awareness of the unit
- Developing relationship mechanisms to ingest records at predetermined times
- Continuing to revisit best practices as holding grow and technologies change

For more information, contact:
NYS Archives Electronic Records Unit
518-474-6926



68

What operational examples have you seen or heard about?

69

Section Five



Goals and strategies for your preservation efforts

Realities

- One size does not fit all
 - Can have different approaches depending on individual record series
- Expect change during planning
 - Changes in resources
 - Retirements, extended leaves, new hires, funding
 - New findings require changes
 - Changes in technologies

71

Fundamental goals

1. Readability of electronically stored information
 - Is the media on which a file is stored still viable?
 - Can the file format be read and displayed properly?



72

Fundamental goals

- 2. Authoritative & trustworthy process
 - Transfer from production to archives
 - House in proper environment
 - Provide secure access and protection



73

Fundamental goals

- 3. Maintain a secure and reliable repository



74

Two preservation strategies needed

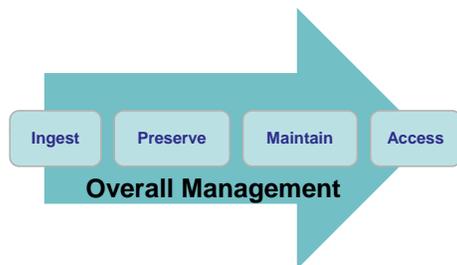
- Process strategy
 - Ingesting, preserving, maintaining, & accessing records
- Marketing strategy
 - Creating awareness of the issues among users & senior officials
 - Ensuring on-going support of preservation efforts

75

Process Strategy

76

Four stages of e-records preservation



77

Overall Management

- Ensuring integrity & authority in process
 - Defined, documented, and consistent processes
 - Mission statement
 - High-level statement describing the purpose of your digital preservation efforts
 - Policy
 - How mission statement will be followed
 - Stated goal & responsibilities; type of custody (legal or physical); roles of personnel; quality control; audit process for compliance



78

Ingest

- You need to determine:
 - What you can and cannot accept
 - File formats, applications, total volume, etc.
 - What is the proper method of transfer
 - Acceptable media/method in which to receive files
 - Process to follow (including an official signoff process)
 - What minimum metadata must be included



79

Ingest

- Metadata
 - Descriptive information that facilitates management of, and access to, the objects being described – “data about the data”
 - Need to maintain metadata as part of complete record to establish authenticity, facilitate retrieval, and to understand the record’s context and relevance



80

Ingest

- Metadata can be:
 - Generated by applications
 - File size, file format, date, time, creator, etc.
 - E-mail metadata: To, From, CC, Subject, Date, Time
 - Hidden: Return path, Unique Message-ID, BCC, etc.
 - Manually entered (or generated from other systems)
 - Classification, subject, title, retention period, record series, additional keywords, etc.



81

Ingest

- Store in a secure repository(ies)
 - Firewall, read only access & intrusion detection
 - Controlled physical access to repository



82

Preserve

- Define content preservation strategy(ies)
 1. Be consistent
 2. Document process followed
 3. Plan for change (as best you can)

83

Preserve - content

- Bit stream copying
 - Multiple instances (replication/copies)
- Refreshing
 - Copying to newer, but same type of media
- System preservation
 - Maintaining hardware, software, media



84

Preserve - content

- Emulation
 - Operating older system environments within new
- Migration
 - Copying from one system or media to another, preserving the essential characteristics of the data (broader than refreshing)
- Normalization
 - Formalized process for conversion to standard formats



85

Preserve - analog alternatives

1. **Printing to archival quality paper**
2. **Computer Output Microfilm (COM)**
 - Converts digital docs to TIFF images
 - Write TIFF images to 16 mm or 35 mm microfilm

Advantages

- Long life expectancy, negates technology obsolescence issues, relatively low cost

Disadvantages

- Suited for static docs, does not address compound documents, web pages, databases, GIS, multimedia, etc.; slow access

86

Maintain

- Quality assurance & control
 - Validate processes
 - Documents procedures, changes over time, & results of QC audits



87

Maintain

- Support
 - Develop formal department level coordinator roles
 - Assist in the identification and transfer of records
 - Do not try and go it alone
 - Infrastructure support
 - Who do you call when something goes wrong
 - Develop Service Level Agreements (SLAs)



88

Access

- Define
 - Who has access & when access is permitted
 - How access occurs
 - Remote; on-site; accessibility limited to documented application, not accessing master copy, etc.
 - Physical security
 - Only authorized personnel; Sign in/out log (date, time & identity)
- Monitor
 - Track who accessed repository
 - What was accessed
 - Potential changes



89

Marketing Strategy

90

Marketing

- Create greater awareness of preservation needs
 - Educate senior administration, users (individuals & depts), legal counsel, risk management and others
- Continuously garner senior level support
 - **Program importance**
 - Infrastructure support (Including financial, IT, & archives staff)
 - Helps keep momentum going



91

But first - know thy audience

1. Need
 - What departments have e-records to preserve?
 - Do you know what's out there?
2. Departmental awareness
 - Do departments know the complexities and potential risks involved in preserving e-records?
3. Trust
 - Do departments believe you have the expertise to protect their files?

92

Not everyone may want to change

- Comfortable with the current process
 - “They are *my* files
 - “*My system* is all I need”
 - “There is nothing wrong with what I am doing!”



93

Typical approaches to “change” (that may not always work)

- Just do it! 
- This dept is onboard, so why aren't you? 
- Do it for the good of the institution! 

94

Potential reaction to new process

- **Early adopters** • Ready to send files or already on board 
- **Bystanders** • Waiting to see how process works with others; Willing to change if early adopters fare well 
- **Resisters** • Defensive, don't want to change; Have other priorities & goals 

95

Change strategies

- **Power** • Use force (aka senior administrators) to change 
- **Rational** • Provide compelling logic or reason (improved protection, access, security, etc.) 
- **Emotional** • Leadership & personal values lead to buy-in; Appeal to need to 'fit in'; risk removal, informal organizational influence 

Source: Chin & Benne 96

Some change strategies to use

- **Early Adopters**
 - Recognition, appreciation, small reward, use as role models
- **Bystanders**
 - Provide inducement to change; provide consistent message; enable to change; do not allow easy escapes/alternatives
- **Resisters**
 - Resistance is futile, get on board or be marginalized (require signoff, send memo stating they agree to assume all risks, cc: senior officials; address early adopters and bystanders first)

Source: Chin & Benne 97

Tell us about...

- Experiences with an 'uncooperative' department or person?
 - How did you handle it?

98

Section Six

Disaster Preparation and Recovery Planning

What constitutes a disaster to you?

- The obvious:
 - 9/11
 - Hurricane Katrina
 - Mount Saint Helens
 - Tsunami
 - Earthquake
 - Flood
 - Fire

100

Don't forget the "small" stuff

- Water pipe break
- Accidental sprinkler discharge
- Gas leak
- Chemical spill from a tanker truck
- Employee sabotage
- Bomb scare
- Computer hardware problem
- Virus
- Data corruption

101

Why is this important to me?

Unique file
 +
Disaster
 =
Lost or damaged records

102

What's needed?

1. Duplicate copy stored at an off-site location
2. Physical access to remote copy when needed
3. Ability to *read* stored copy of record
 - Same backup drive & other hardware at off-site?
 - Same version backup software, operating system and device drivers to restore the files?
4. Documented policies & procedures
5. Designate who is leading preparedness effort
6. Awareness education & procedural training
7. Periodic testing

103

What not to do

- Store backup next to original
- Assume the backup worked as intended



Daily backup tape sitting next to server

104

Steps to prepare



1. Does a plan exist already?
 - Does the plan fit *your* environment?
 - Modify an existing plan to meet your needs
2. Perform a disaster preparedness assessment
 - Identify the valuable, single source records
 - Identify proprietary, non supported or failing systems
3. Develop/modify disaster recovery plan
4. Train staff on procedures
 - Include prevention steps & general awareness
5. Testing & retesting of plan (continuous refinement)

105

Things to consider

- Regional collaboration
 - Leverage other site locations as a hot site (e.g. local governments, businesses, other institutions)
 - Sharing mutual supplies
- Establish mutual aid agreements
- Coordinating with other emergency plans
 - Emergency Operations Plan (EOP)
 - Continuity of Operations Plan (COOP)
 - IT Backup & Disaster Recovery planning
 - Building evacuation plans

106

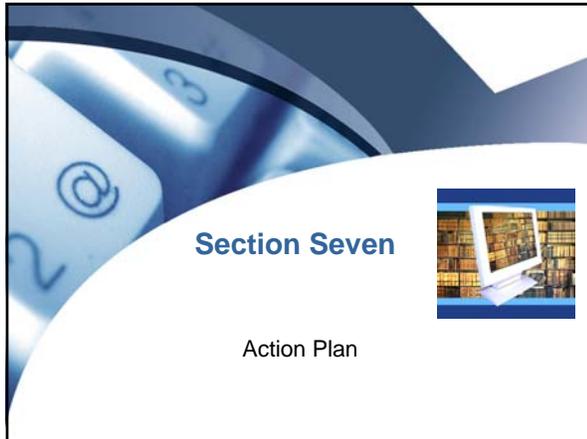
Have you experienced any disasters?

107

Additional resources

- Society of American Archivists
 - <http://www.archivists.org/mayday/>
- Council of State Archivists (COSA)
 - Intergovernmental Preparedness for Essential Records (IPER)
 - www.statearchivists.org
- NYS Historical Records Advisory Board
 - <http://www.nyshrab.org/disaster/index.shtml>

108



Initial goals

- Focus on target area (start small)
- Establish and document a systematic process for the transfer of the records to the archives
- Plan for change



110

Realities

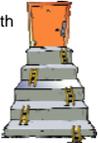
- Can't do everything
- Limited budget



111

Digital preservation steps

1. Planning
 - Develop *initial* strategy for program
 - Create awareness & gather support
 - Identify target areas (may change as you progress)
2. Inventory
 - Areas you worked with
 - What other areas should you work with
3. Analysis
 - Examine discovered data, research best practices & reconfirm support



112

Digital preservation steps

4. Design
 - Infrastructure (IT system & process)
5. Implementation plan
 - Pilot group & target record series
6. On-going activities
 - Awareness (users & senior administration)
 - Quality assurance processes
 - Refine processes & tools
 - Education (archives and users)



113

Planning phase

- Develop initial strategy for program
 - Will be refined later on
 - Develop mission statement
- Identify targeted areas
 - List all functional areas within the institution
 - Identify contacts in each area
 - Don't try it all; start with existing users and relationships
 - Pick 1-2 areas to start
- Find partners & collaborators
 - Sounding board; outside advice from other institutions, local governments, experts



114

Planning phase



- Create awareness
 - Among users & senior officials
 - Workshops, flyers, e-mail, get senior admin to discuss at staff meetings; hold informal 1-on-1 meetings, etc.
- Gathering support
 - From audience who approved the initiative
 - Identify potential support help group (e.g. IT, legal, etc.)
- Gather data to back up the need
 - Information gathering and preparing the facts
 - Potential issues, past examples (internal and at other institutions)
 - Continuous process

115

Inventory phase

- Determine scope
- Define departments & other functional areas
- Identify key contacts
- Start with 1-2 departments



116

Inventory phase

1. Know what e-files exist
 - Which are deemed records?
 - What are essential records?
 - What have long-term retention requirements (e.g. greater than 10 years)?
 - What items have historical or other intrinsic value?
 - Use retention schedule to assist in identifying
2. Identify ownership



117

Inventory phase

3. Know where files are stored & file formats
4. Know the output (official copy)
 - Get general idea of where duplicate items go (also helps in determining disaster assessment risk)
5. Identify generating systems & applications



118

Inventory phase

- Understand high-level work process flow
 - Who creates records? (different from owner?)
 - How do files arrive in office?
 - How they are used in office? (Who gets a copy?)
 - How many different renditions exist?
 - Where are files currently stored?

Understand use before designing filing plan or file transfer process



119

Analysis phase

- Inventory analysis
 - Retention requirements
 - Preservation needs
 - Proprietary systems (including in-house developed)
- Research
 - Best practices
 - Peer institution discussions & review
- Determining future access needs
 - Narrowing the target records needing more frequent access



120

Analysis phase



- Establish & document roles & responsibilities
 - Who is the 'official' record owner?
 - Who maintains the generating applications?
 - (e.g. application administrator for financial system, etc.)
 - Who is in charge of the infrastructure
 - Network issues and reliability, technical advisory services
 - Server-based directory/folder administration
 - Who could be Preservation Coordinators?
 - Focal point within each functional area for administration, guidance, etc.

121

Analysis phase



- Develop proper metadata
 - Data describing content, context and structure of e-records and their management over time
 - How to interpret the bit stream?
 - What are the generating applications & system information?
 - How do we provide proper classification and organization?
 - How to prove easy access? (e.g. searching & retrieving)
 - How best to provide the records in proper context?
 - What can be system generated vs. manually created?

122

Design phase



- Based on what you know now:
 - Review mission, goals and initial strategy with senior administration and concerned parties
 - Discuss with peer institutions, IT, Legal, key departments, senior administration; research best practices
- Do not be afraid to change course
 - Is the scope too large or too small?
 - Do we start with another department for pilot?

123

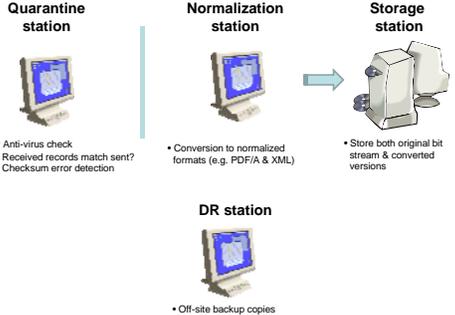
Design phase



- Determine software tool set
 - Open source software
 - Commercial software
 - Develop your own
 - Important – know your support capabilities!
- Determine infrastructure
 - Storage environment
 - Physical space (shared or dedicated)
 - Storage repository (media and accepted formats)

124

Design phase



Quarantine station



- Anti-virus check
- Received records match sent?
- Checksum error detection

Normalization station



- Conversion to normalized formats (e.g. PDF/A & XML)

Storage station



- Store both original bit stream & converted versions

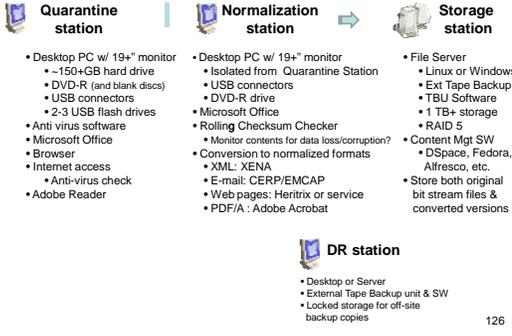
DR station



- Off-site backup copies

125

Design example #1



Quarantine station



- Desktop PC w/ 19"+ monitor
- ~150+GB hard drive
- DVD-R (and blank discs)
- USB connectors
- 2-3 USB flash drives
- Anti virus software
- Microsoft Office
- Browser
- Internet access
- Anti-virus check
- Adobe Reader

Normalization station



- Desktop PC w/ 19"+ monitor
- Isolated from Quarantine Station
- USB connectors
- DVD-R drive
- Microsoft Office
- Rolling Checksum Checker
- Monitor contents for data loss/corruption?
- Conversion to normalized formats
- XML: XENA
- E-mail: CERP/EMCAP
- Web pages: Heritrix or service
- PDF/A : Adobe Acrobat

Storage station



- File Server
- Linux or Windows
- Ext Tape Backup
- TBU Software
- 1 TB+ storage
- RAID 5
- Content Mgt SW
- DSpace, Fedora, Alfresco, etc.
- Store both original bit stream files & converted versions

DR station



- Desktop or Server
- External Tape Backup unit & SW
- Locked storage for off-site backup copies

126

Design example #2 – lower cost

<p>Quarantine station</p> <ul style="list-style-type: none"> • Desktop PC w/ 17" monitor • ~100+GB hard drive • DVD-R (and blank discs) • USB connectors • 2-3 USB flash drives • Anti virus software • Microsoft Office • Browser • Internet access • Anti-virus check • Adobe Reader 		<p>Normalization station</p> <ul style="list-style-type: none"> • Desktop PC w/ 17+ monitor • Isolated from Quarantine Station • USB connectors • DVD-R drive • Microsoft Office • Conversion to normalized formats • PDF/A: Adobe Acrobat, Microsoft Office add-on to PDF/A, Open source (PDFCreator) • E-mail: Batch print to PDF/A • Web pages: Use 3rd party service; print static pages to PDF/A 	<p>➔</p> <p>Storage station</p> <ul style="list-style-type: none"> • 1 TB External HD(s) • Store files in native Windows directory structure • Store both original bit stream files & converted versions
<p>DR station</p> <ul style="list-style-type: none"> • Desktop PC (optional) • 1 TB External Hard Drive(s) (test regularly) • Locked storage for off-site backup copies 			

127

Design phase

- Document process
 - Acceptable file formats, applications, total volume, etc.
 - Acceptable media in which to receive files
 - File transfer/custody process (including an official sign off process)
 - File acceptance review, verify authenticity (e.g. use Cyclic Redundancy Check (CRCs) codes to verify)
 - Minimum metadata to include
- Develop basic policies & procedures
 - Be consistent & concise; periodic review & auditing
- Define access & protection



128

Implementation phase

- Scope and objectives defined
 - Target area identified
 - Initial record series identified
 - Preservation Coordinator(s) identified
- Preliminary pilot project schedule developed
 - Test process with pilot group
 - Review process
 - Refine as needed
 - Plan for next test group or record series

129

Implementation phase

- Disaster preparedness
 - More than just backing up a system
 - What steps do we take if a disaster occurs?
 - Written, tested, an updated annually
 - Communicate & train staff on plan



130

On-going phases

- Remember, the work is only beginning...
 - You only scratched the surface
 - Once you think you figured things out, they change
 - Other priorities come up and attention is diverted



131

On-going phases

- Awareness (users and senior administration)
 - Constant reinforcement: discuss at staff meetings & higher profile meetings, hold periodic training/workshops
 - Incorporate policy & procedures into training
- Review processes
- Refinement processes & tools
- Education (both archives' staff and users)



132

Exercise

- You get a call from a department:
 - “I have a bunch a files that need to keep intact, but I do not know what to do with them. Can I give them to you?”
- Write down a list of questions you should ask the person (and yourself) about the files.

133



Section Eight

Wrap up



Goal of preservation

- Ensuring the use and accessibility of information in a record within the proper context for that record's full retention period



135

Fundamental shift for archivists

- Need to know more about record creation
- Need to develop greater number of partnerships
- Need to learn more about key technologies
- Must rely on others for assistance



136

Realities

- Never will have all the needed resources
 - This still hasn't changed...
- Expect same or worse issues as paper
 - If employees did a poor job managing paper, expect the same with digital records
- Get cooperation and buy-in
 - Stress business case for preserving digital records
- No single best strategy
 - Often multiple strategies depending on situation

137

Digital preservation fundamentals

1. Readability of e-stored information
 - Proper media, file format, hardware & software
2. Authoritative & trustworthy process
 - Transfer, management, protection, access
3. Maintain a secure and reliable repository
 - Trusted digital repository



138

What we discussed

- Technology is only one part of the challenge
 - Organizational, legal, available resources, etc.
- Greater reliance in e-records
 - More are born and remain digitally today
- **Creating awareness of the issues**
 - Senior officials, users, archivists, & records mgrs
- What affects readability?
 - Media, formats, devices, software, hardware



139

What we discussed

- Be aware of existing standards & developments
- Have you performed a disaster preparedness assessment?
- Is there a (tested) recovery plan in place?
- Know the steps to follow...



140

Digital preservation phases

1. Planning (gather support, develop strategy, target areas)
2. Inventorying
3. Analysis
4. Design (infrastructure & procedures)
5. Implementation plan (pilot group & record series)
6. On-going activities (awareness, education, refinement)



141

Best (better) practices



1. Develop good e-records management habits

- Manage electronically & centrally
- Implement schedules & standards at creation
- Develop proper classification system
- Simplify retention as much as possible
- Destroy records when appropriate
- Develop overall policy incorporating e-records
- Form partnerships with key admin groups & IT
- Strategies apply beyond just *permanent* e-records



143

2. Be proactive

- Find out what is out there
 - Vulnerable records; current and planned systems & applications, users, etc.
- Be aware of standards and best practices
- **Leverage & coordinate resources**
- Document policies & procedures (be consistent)
- Continuously create awareness
 - Educate senior administrators & users
- Perform periodic review & auditing



144

