e-Records Conference 2006

To Preserve and To Protect

A Conference for Texas State Government and Universities

Sponsored by the
Texas State Library and Archives Commission (TSLAC)
and the
Texas Department of Information Resources (DIR)

Texas State Library and Archives Commission

DIR Department of Information Resources

www.tsl.state.tx.us/slrm/conferences/erecords.html

Conference Agenda

Registration and Exhibit Area Open	8:00-8:30 a.m.
Welcoming Remarks DIR Speaker Dr. Michael Heskett, State Records Administrator	8:30-8:45 a.m.
Featured Speaker Charles Dollar, Senior Consultant, Cohasset Associates	8:45 a.m.-Noon
Lunch (provided) & Exhibit Area Open	Noon-1:00 p.m.
Afternoon Session Dean McKinney, Vice President, BMS CAT	1:00-4:15 p.m.

The exhibit area will be open during registration, the mid-session breaks, and lunch. Please visit the exhibits to learn more about related products and services.

Welcome from TSLAC and DIR

The Texas Department of Information Resources and the Texas State Library and Archives Commission welcome you to the e-Records Conference 2006.

Over the years, we have provided these conferences for Information Resources Managers, Records Management Officers, and other Texas government staff who have the common goal of managing government information, but an uncommon approach to accomplishing this daunting task. We bring together an extraordinary group of diverse government staff to listen, learn, evaluate, connect and return to their offices ready to collaborate in effective ways to gain control over an ever-growing volume of government data.

When establishing and managing information management systems, we must consider retention requirements; the Public Information Act; privacy laws; audit standards; legal requirements; historical value; staff needs; technical systems and standards; a complex legislative environment; and the service requirements and expectations of the residents of Texas. It is in finding common ground that we find uncommon success.

This year's conference focuses on the preservation and protection of state electronic records. Charles Dollar, a leading expert in digital preservation will discuss issues surrounding portability and persistence of digital objects as well as guidelines for selecting file formats for long term retention. Dean McKinney, Vice President with BMS CAT, a worldwide disaster recovery company, will examine prevention and recovery practices relating to electronic media and facilities that house the media.

It is with great pleasure that we welcome each of the vendors that are here today. They are an integral component of the partnerships enabling Texas government organizations to move forward in the transition from paper to pixels. They stand on common ground with every level of government in Texas in their desire to implement more effective systems for information management and access. Please take advantage of the opportunities to visit with the vendors who join us here today.

We are so glad you are here. We look forward to bringing you an informative program and the opportunity to connect with other individuals seeking to manage electronic records legally and effectively.

Welcome to the e-Records Conference 2006!

Texas State Library and Archives Commission

Department of Information Resources

Agency and Speaker Notes

Sponsoring Organizations

DIR

The Texas Department of Information Resources (DIR) is charged with providing leadership and coordination of information resources for state agencies and universities. Toward this goal, DIR is committed to developing and implementing innovative management strategies for the efficient and accountable use of information technology in state government.

DIR is responsible for developing and managing a vision of IT-enabled government, establishing rules and standards for information technology, negotiating volume purchase agreements for quality products and services, and providing telecommunications services to agencies and local governments, among others. To read more about DIR, visit our Web site: www.dir.state.tx.us

TSLAC

The Texas State Library and Archives Commission (TSLAC) is responsible for providing guidance and recommendations to uphold select legislative mandates. The division charged with coordinating the e-Records conference is State and Local Records Management (SLRM). SLRM assists Texas governments in establishing and implementing records management programs. Serving exclusively state agencies and local governments, the division offers training classes, consulting services and forms needed for all aspects of records and information management. Whatever the situation, SLRM staff are always close at hand, ready to help. To find out how SLRM can help you, visit our Web site: www.tsl.state.tx.us

Featured Speaker

Charles Dollar

Charles M. Dollar is an internationally recognized expert on the life cycle management of electronic records, particularly digital preservation. His experience includes:

- Member of the History Department Faculty at Oklahoma State University (1966-1974)
- Twenty-years of experience in working with electronic records at the National Archives and Records Administration (1974-1994)

- Archival educator at the University of British Columbia (1994-1999)

A Senior Consultant at Cohasset Associates since 1999, Mr. Dollar is a frequent speaker on electronic records management issues at conferences and workshops in the United States and around the world and the author of more than twenty publications and reports on life cycle electronic records management, with a focus on digital preservation.

One of his books, *The Impact of Information Technologies on Archival Principles and Methods* (1992), which was translated in German, Italian, and Croatian, received a special commendation from the Society of American Archivists in 1993. His most recent book, *Authentic Electronic Records: Long-Term Access Stragegies* (1999) received the 2000 Waldo Gifford Leland award from the Society of American Archivists "as a superior monograph on archival theory and practice." He currently is preparing a second edition of this book.

Mr. Dollar, who holds a Ph.D. in History from the University of Kentucky, is a Fellow of the Society of American Archivists. He is the past Chair of the Association of Information and Management Association (AIIM) Standards Board, which has the final review of all standards that require the approval of AIIM. He is the 2005 recipient of the Emmett J. Leahy Award for "His Outstanding Contributions to the Information and Records Management Profession."

Dean McKinney

A Florida native, Dean McKinney joined the Blackmon-Mooring Steamatic of Fort Worth, Texas as a commercial sales representative. A top producer in that operation, Dean then assumed a position with the Fire and Water Damage Division prior to accepting a position in Denver, Colorado as general manager of company operations in the Rocky Mountain Region.

Due to his success in establishing himself as a regional leader in the Rocky Mountain area, Dean was offered the position of regional manager of the Northeastern United States for BMS Catastrophe with offices near Washington, D.C. In 1997, Dean was named vice president of BMS CAT.

During his tenure with BMS CAT, Dean has been responsible for some of the company's largest projects. Among these projects are the three largest records center recoveries in history. Dean McKinney is internationally recognized as an industry leader in the field of fire and water damage recovery but, in particular, he has become an authority on records center, archive and library recovery and is also a frequent speaker on the subject. More recently, Dean directed recovery efforts of the U.S. Pentagon Library and Archives as a result of damage sustained in the terrorist attacks of September 11th. Also, the recovery of the Jamestown Museum and Archives in Williamsburg, Virginia after damage from Hurricane Isabel. Most currently he handled flood recovery at the National Archives main facility on the mall in Washington, D.C., and the Washington National Records Center in nearby Maryland.

Today Dean McKinney is based in the corporate offices of BMS CAT in Fort Worth, Texas. His responsibilities are marketing and direction of the entire records, book and media recovery division, as well as special collections, museums and historic buildings.

Dean has been married for more than 25 years and he and his wife have three college-aged children. Even with a schedule that is oftentimes hectic, his top priority is his family. Although his clients consume a great deal of his time, he stresses that professional success must be balanced with one's personal priorities of family, church, and community.

Exhibitor Contact Information

October 19, 2006

TSLAC and DIR wish to thank exhibitors for participating in the 2006 e-Records Conference. Attendees are invited to visit the exhibits during the breaks.

ARMA Austin Chapter
www.arma-austin.org
Contact: Stephanie Jones
(512) 305-9193
Stephanie.jones@agr.state.tx.us

CA, Inc.
www.ca.com
Contact: Brooke Davis
(404) 946-1080
Brooke.davis@ca.com

FileNet Corporation
www.filenet.com
Contact: Dennis Lowe
(512) 502-1770
dlowe@filenet.com

The Harding Group, Inc.
www.harding-group.com
Contact: Vickie Key
(817) 461-3393
v@harding-group.com

MTM Technologies, Inc.
www.mtm.com
Contact: Brad Truman
(512) 469-3510
btruman@mtm.com

Oracle
www.oracle.com
Contact: Amanda Campbell
(512) 329-5021
amanda.campbell@oracle.com

TexasOnline
www.texasonline.com
Contact: Chris Keel
(512) 542-8043
chris.keel@bearingpoint.com

Brother International Corporation
www.brother.com
Contact: Jim Sharkey
(908) 252-3089
Jim.sharkey@brother.com

CGI-AMS
www.cgi-ams.com
Contact: Cheryl Morano
(512) 493-3529
Cheryl.morano@cgi.com

IBM
www.ibm.com/software
Contact: Rene Brito
(512) 469-6504
rbrito@us.ibm.com

Neubus, Inc.
www.neubus.com
Contact: Eric Kennedy
(512) 833-6627
ekennedy@neubus.com

PacoTech, Inc.
www.pacotech.com
Contact: Gayle R. Page
(713) 688-0404
gpage@pacotech.com

Feature Presentation

Part 1:

Preserving Digital Records: A Tale of Bits and Bytes

Charles Dollar, Cohasset Associates

Preserving Digital Records

All About Bits and Bytes

Texas State Library and Archives Commission

E-Records Conference 2006

October 19, 2006

Charles Dollar

Session Overview

- Challenge of digital records preservation
- Digital records preservation approaches
- Digital Records Formats and Legal Admissibility
- Digital Records Formats and Preservation
- Looking to the Future

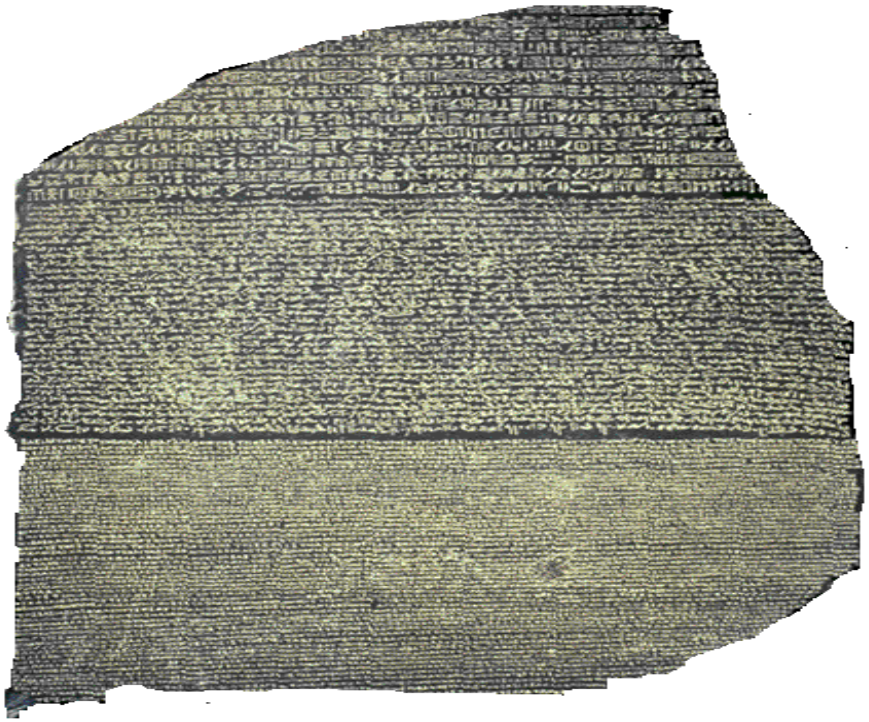
Part 1
Challenge of Digital Preservation

Digital Preservation Issues

1. Mind set - digital v. traditional information artifacts
2. Digital records must be stored and retrieved digitally
3. Digital records require an “interpreter”
4. Hardware/software dependence
5. Technology obsolescence

Digital Representation of Information Objects

Rosetta Stone (196 BC)



Rosetta Stone Attributes From A Modern Records Perspective

- Physical
- Logical
- Conceptual

Digital Object Attributes

- Physical
- Logical
- Conceptual

Physical Attributes of Digital Records

- Signals on a storage medium
- Binary representation
- No inherent intelligence or meaning – a string of representations of 1s and 0s

What Do Binary 1s and 0s Represent?

Decimal

"50"

Number

"2"

0110110010101110000001100100000110000100010011111011

Vector Drawing

Photograph

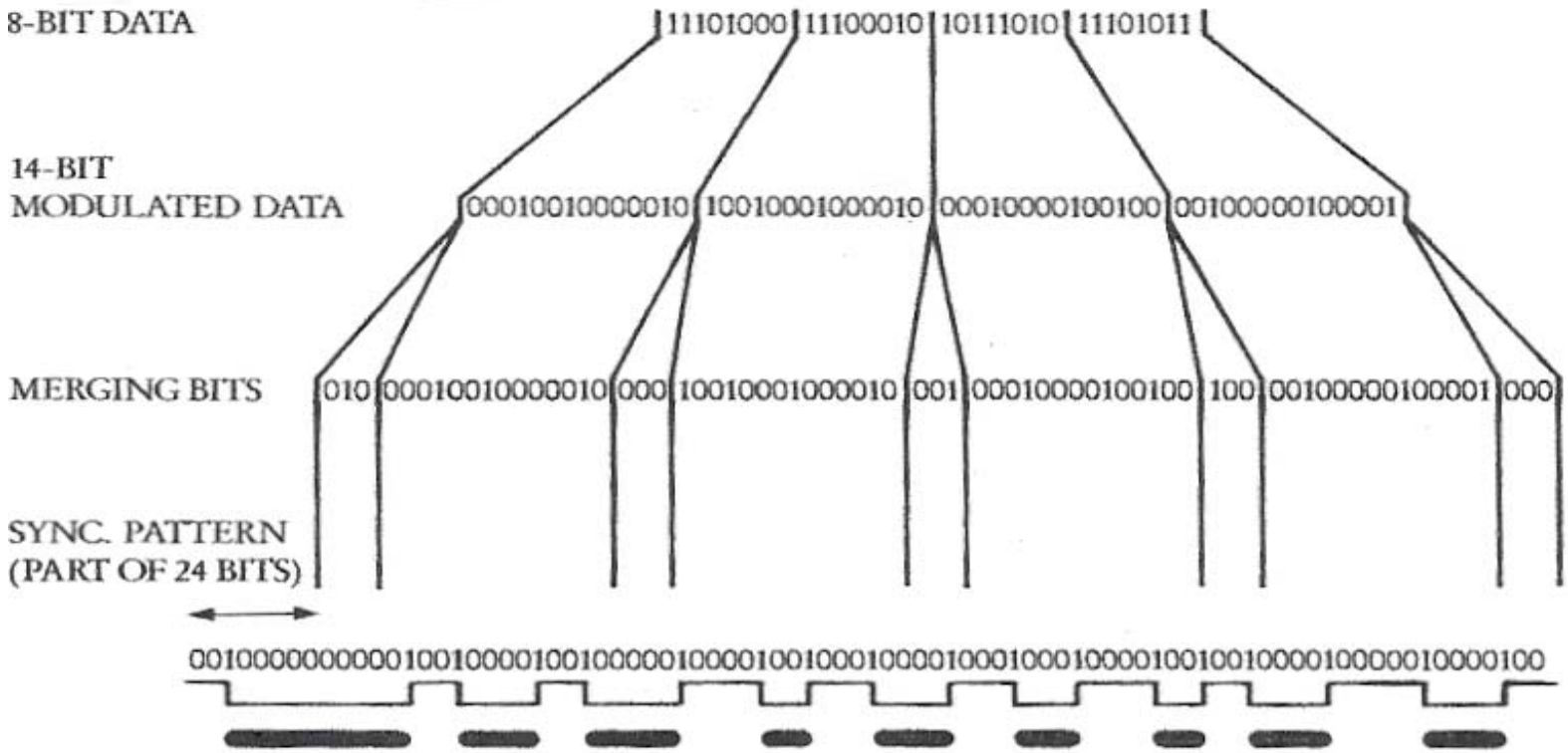
Spreadsheet

Sophie

TIFF

JPEG

CD-ROM Data Encoding



Logical Constructs

- Software applications recognize a logical object based on data type
- Software applications recognize file formats ASCII, native word processing, PDF, PDF/A, TIFF, JPEG
- Software applications interpret bit streams and render them into human readable form

What is this?

0111011101111000011110010011000 1

0011001000110011001101000011010 1

0111000000110010010000010011010 1

0011001100110101001100010011010 1

0011010000110111001110000011100 1

0011100000110111001100000011000 0

0011001000110011001101000011010 0

A Clinical Trial Data Record

0111011101111000011110010011000 1
0011001000110011001101000011010 1
0111000000110010010000010011010 1
0011001100110101001100010011010 1
0011010000110111001110000011100 1
0011100000110111001100000011000 0
0011001000110011001101000011010 0

XYZ12345P2A535I5478987002344000215679 9

A Clinical Trial Data Record
Conceptual Structure – Context & Metadata

XYZ1234 I5478987002344000215679 9

	Field Definitions (digital)
Project No.	XYZ12345
Protocol	P2A5351
Investigator	47897700
Patient ID	23440
Visit No.	002
Result	1576799

A Clinical Trial Data Record

Conceptual Structure –Context & Metadata

XYZ12345P2A535I5478987002344000215679 9

	Field Definitions (digital)	Code Rules Definitions
Project No.	XYX12345	Research
Protocol	P2A5351	Happy Daze
Investigator	47897700	Dr. Nuzzle
Patient ID	23440	
Visit No.	002	Cumulative
Result	1576799	Euphoric

A Clinical Trial Data Record

Conceptual Structure –Context & Metadata

XYZ12345P2A535I5478987002344000215679 9

	Field Definitions (digital)	Code Rules Definitions	Derived Values (digital)
Project No.	XYX12345	Research	
Protocol	P2A5351	Happy Daze	
Investigator	47897700	Dr. Nuzzle	
Patient ID	23440		
Visit No.	002	Cumulative	
Result	1576799	Euphoric	

Journey of a Byte

Creation/Storage

- Key “A”
- Converted to ASCII Decimal 65 (01000001)
- Saved as Doc
- Operating system notifies drive controller
- Drive controller receive 01000001
- Recodes to 1100011100
- Writes to storage media

Retrieval/Display

- Application sends request to disk controller
- Drive controller notifies drive to find “A” and copy it
- Drive strips off recoded values to original 65 (01000001)
- Notifies Operating System to pick up the “A”
- Operating system delivers 0100001 to the App. System
- App. recognizes ASCII text format
- Renders “A” for viewing

Take Away

- Electronic records exist only when they are rendered
- Preserve the capacity to render the representations of 1s and 0s into a human understandable form
- Ensure the trustworthiness rendered records

Part 2

Four Digital Preservation Approaches

The Museum Approach

- Save old hardware and software keep them operational
- Garage computer hobbyists
- Short term solution
- Presumes that people in the future will want to use old hardware and software
- Requires that the bit stream is kept alive

Data Archaeology Approach

- Seamus Ross and University of Glasgow
- Minimalist strategy
  - Ensure the continuing readability of digital records
  - Collect documentation about software - operating systems and applications
  - Convert to new technology on demand (emulation on demand/reverse engineering)

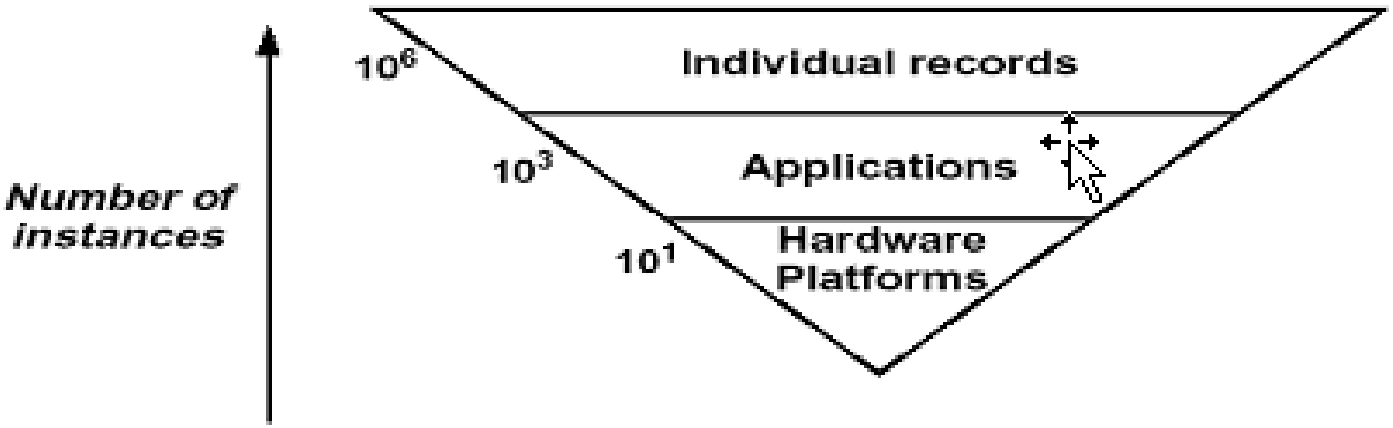
Emulation Approach

- Jeff Rothenberg, Rand Corporation

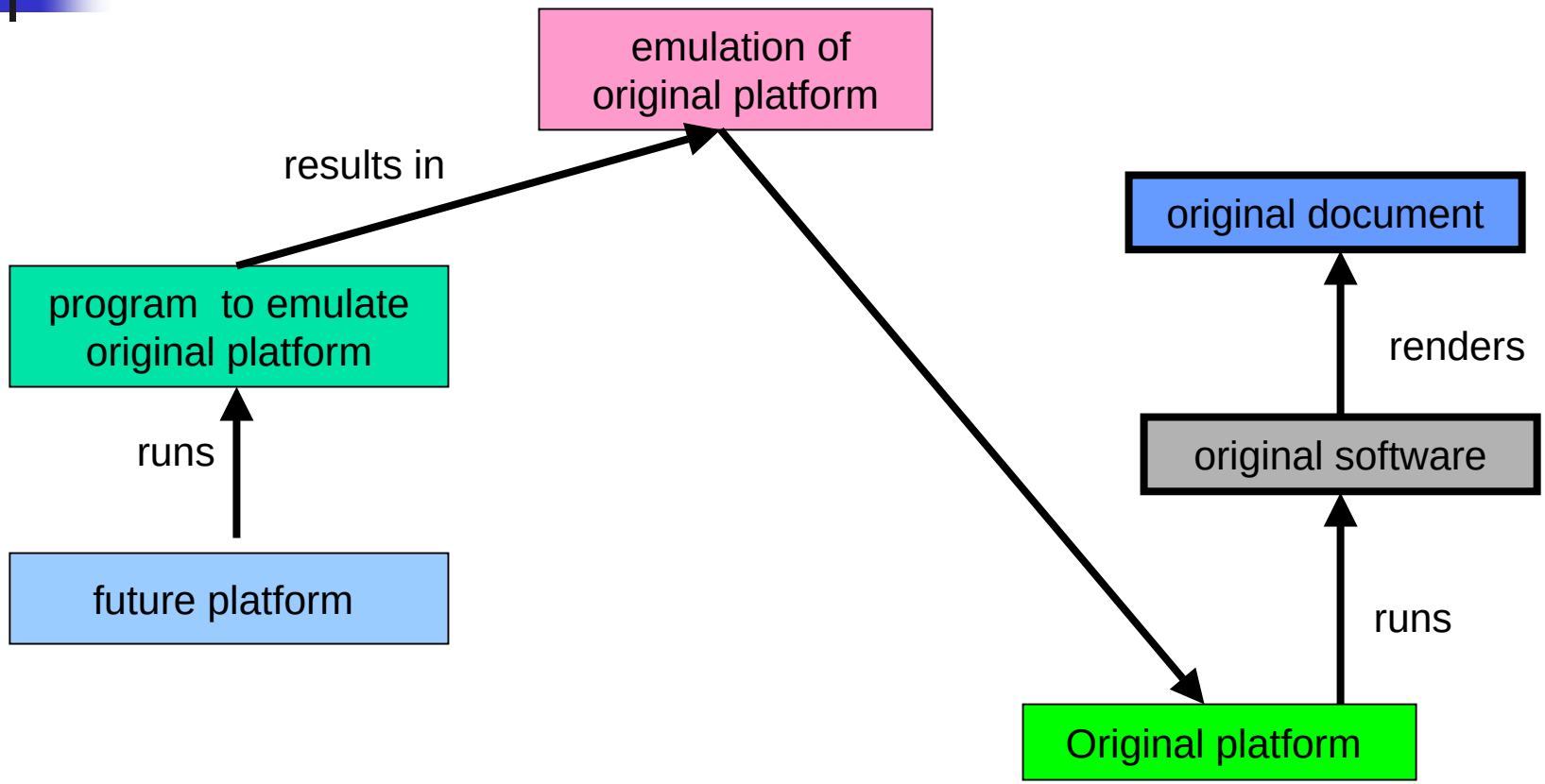
  "Emulation is a process in which one computer is used to reproduce the behaviour of another computer with such fidelity that the emulation can be used in place of the original computer"
- Supports executable "digital originals"
- Requires native application and emulator of original computer platform
- Presumes on-going readability of digital records

Why Emulate Hardware

- Hardware is easier to describe (i.e., specify) than software
  - Since it must be manufactured
- Hardware is more stable than software
  - Since it is harder to change or upgrade
- There is a lot less hardware to emulate than software
  - Since it is harder to change or upgrade



Overview of Emulation



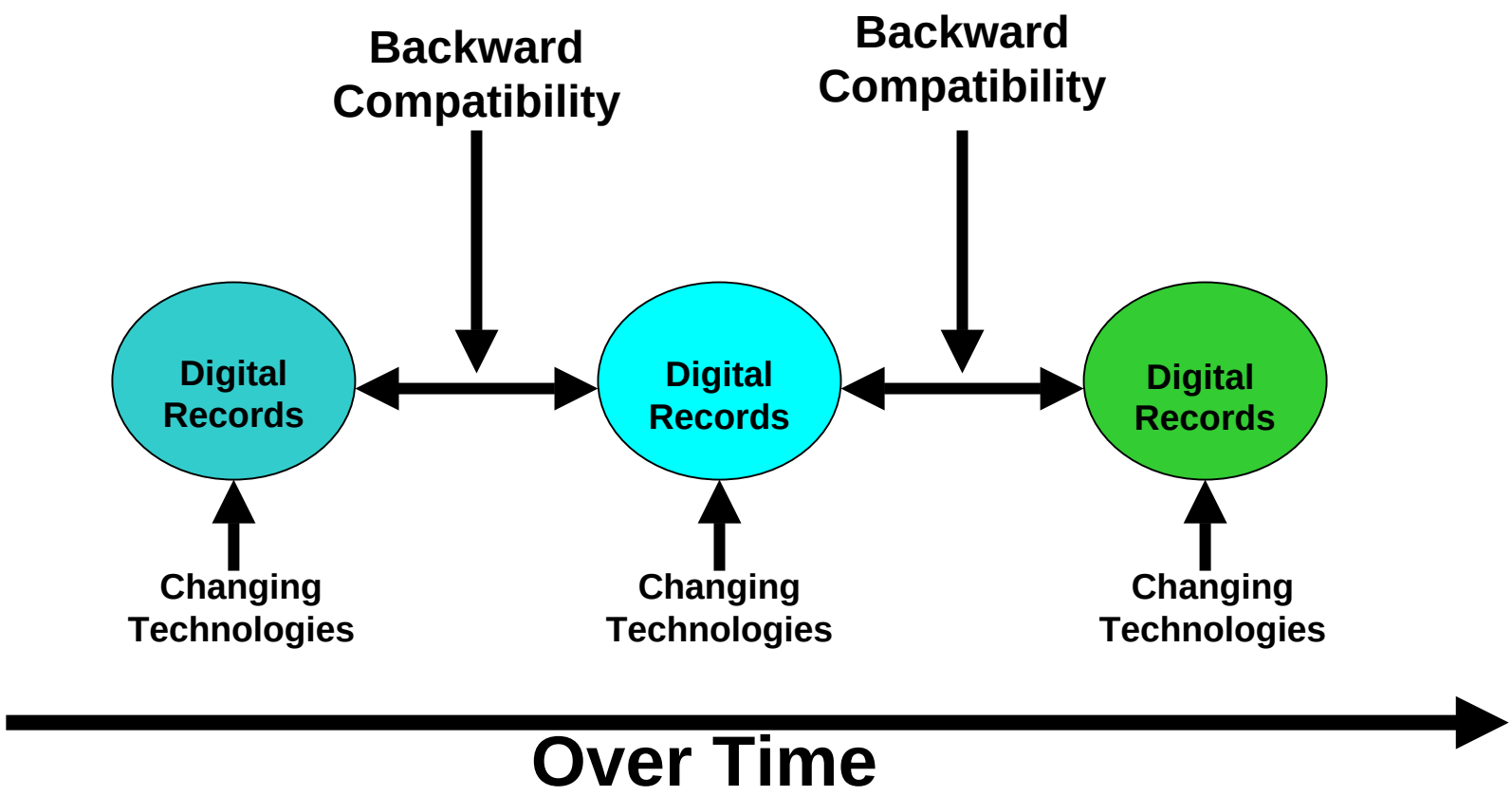
Emulation Problems

- Users will have to know how to run obsolete software
- Likely to require vernacular copies
- Bit streams must not change
- May have to emulate more than processors
- No specifics yet on a commercially viable emulator specification

The Migration Approach

- Ensure usable and trustworthy digital records for as long as necessary without regard for computer technology platform
- Presumes “readable” digital records
- Involves converting digital records to technology neutral file formats
- Requires backward compatibility
- Preserves processibility
- Potential risk

Conversion Life Cycle Process



Legacy Records Migration Overview

- What constitutes legacy records
  - Software/operating system dependent
  - No export functionality
  - Mapping requires extensive coding
- Brodie & Stonebreaker, *Migrating Legacy Systems* (1995)
  - Ten steps
  - Labor intensive
  - Costly
  - Possibility of failure

Migration Technology

- Analyze legacy information system
- Decompose the legacy system structure
- Design the target interfaces
- Design the target applications
- Design the target database
- Install and test target environment
- Create necessary gateways
- Migrate the legacy records
- Migrate the legacy application
- Migrate the legacy interfaces

The Pros and Cons of Migration

Pro

- Methodology is understood and demonstrated – music transcription
- Utilizes today's technologies
- Builds on "backward compatibility" where possible
- Employs open standards
- Adaptable to many digital records

Con

- Standards are not a panacea
- Digital record bit stream changes
- Potential to introduce errors without quality control
- Complex inactive digital records migration is costly
- May lose some of original look and feel
- Never ending

Take Away

QUIZZ

What is it about migration that distinguishes it from the other three digital preservation approaches?

Part 3
Digital Records File Formats and Legal Admissibility

What File Formats Do

- Tell operating system what is the record type
- Tell application software how to interpret 1s and 0s
- Specify the internal logical arrangement of digital objects
- Provide special instructions - compression algorithms
- Provide information understood by specific application software

What Is Legal Admissibility?

- Admitting of records as evidence in regulatory or legal proceedings, including discovery
- What does this have to do with digital preservation?
- Trustworthiness

Legal Admissibility Requirements

Federal Rules of Evidence, Federal Business Records Act, Federal Uniform Photographic Copies of Business and Public Records as Evidence Act

- Accuracy – creation & reproduction
- Integrity – unaltered
- Reliability – repeated and consistent manner in the ordinary course of business
- Readability – readable as original
- Accuracy of reproduction
- Trustworthiness – what a record purports to be

Types of File Formats

- Text capture
- Text reproduction
- Vector graphics capture
- Vector graphics reproduction
- Original bit map images
- Reproduction bit map images
- Reproduction of sound recording

Original Text Capture

- ASCII
- Native word processing
- Marked-up

Reproduced Text

- Plain ASCII
- Rich Text Format
- HTML
- XML
- PDF
- PDF/A

Brief Portable Document Format Tutorial

- Device and software independent
  - Character and vector data
  - Absolute fidelity of rendered documents
  - Versatile
- How PDF documents are created
  - Postscript background
  - Postscript page matrix (612/792)
  - Mathematically defined

Brief Portable Document Format Tutorial (2)

Save as PDF

- Acrobat Distiller
- Temporary Postscript – fonts embedded
- Distiller converts Postscript to PDF
- Distiller settings
  - Encryption
  - Security - who can view, print, extract and edit
- Reduced storage

Render PDF

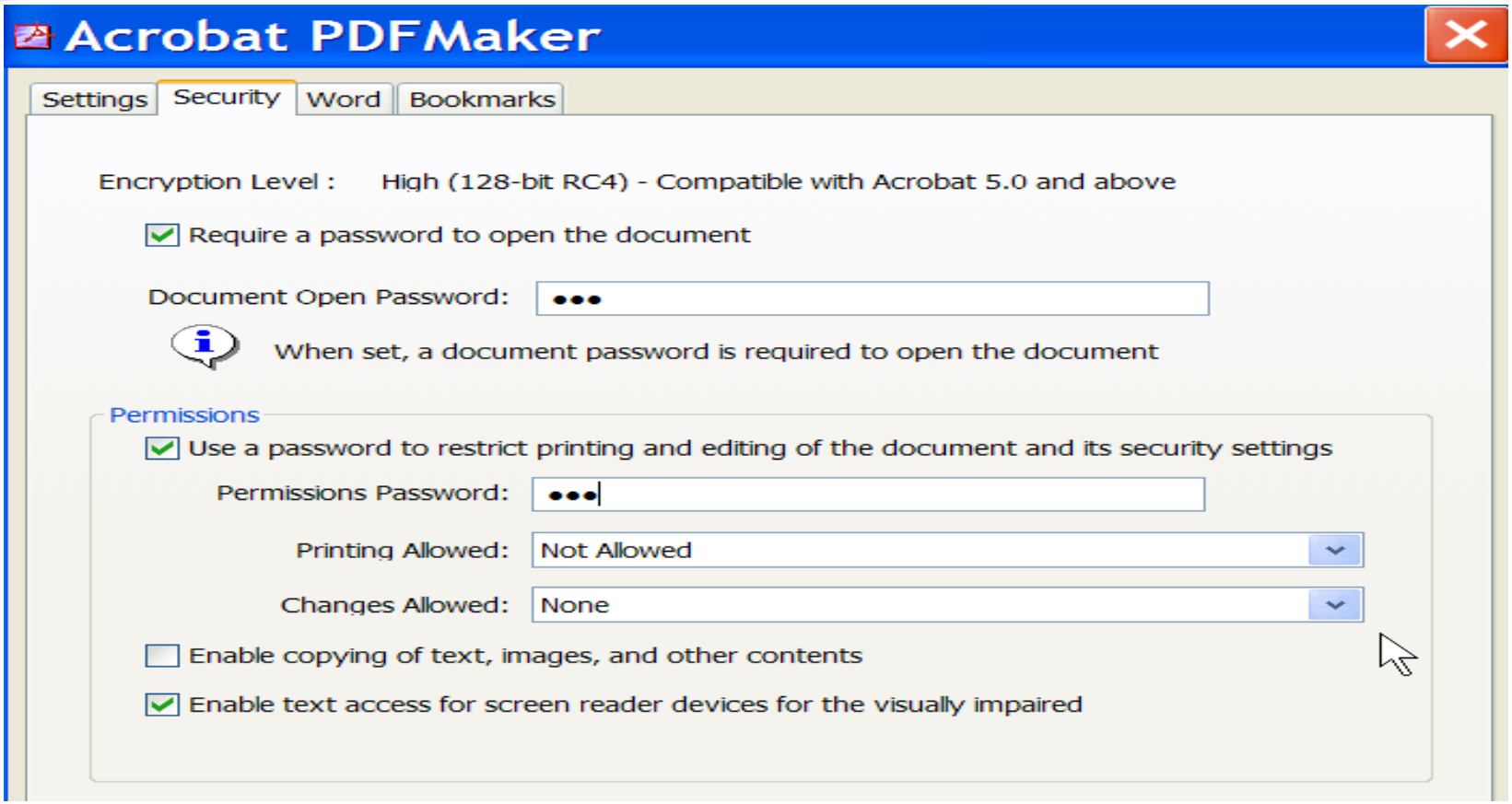
- Sent to Postscript
- Converted to temporary Postscript document
- Rasterized into a bitmap (standard procedure)
- Rendered – screen or printer

What A PDF Document Looks Like


```
%%EOF

39 0 obj<</Length 95/C 97/Filter/FlateDecode/I 119/L 81/S 49>>stream
xÚb```b``î□“Ÿ□¸□□€□(ÆÄÀÃÀqà»□Ã□¦□ □□□□X¡˜□Á‡□[@-$˜□ï□ˆÇÃÀ :“□l0ƒ□□330□k□iF P>\7□□ƒº+D”‘□ À
*□□H
endstream□endobj□38 0 obj<</Length 20/Filter/FlateDecode/W[1 1 1]/Index[9
22]/DecodeParms<</Columns 3/Predictor 12>>/Size 31/Type/XRef>>stream
xÚbbbf`b``¤6□□0 □M G
endstream□endobj□32 0 obj<</MarkInfo<</LetterspaceFlags 0/Marked true>>/Metadata 7 0
R/PieceInfo<</MarkedPDF<</LastModified(D:20060819120004)>>>>/Pages 6 0
R/PageLayout/OneColumn/StructTreeRoot 9 0
R/Type/Catalog/Lang(EN-US)/LastModified(D:20060819120004)/PageLabels 4 0 R>>□endobj□33 0
obj<</CropBox[0 0 612 792]/Parent 6 0 R/StructParents 0/Contents 37 0 R/Rotate 0/MediaBox[0
0 612 792]/Resources<</Font<</TT0 34 0 R>>/ProcSet[/PDF/Text]/ExtGState<</GS0 36 0
R>>>>/Type/Page>>□endobj□34 0 obj<</Subtype/TrueType/FontDescriptor 35 0 R/LastChar
148/Widths[250 0 0 0 0 0 0 0 333 333 0 0 250 0 250 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 722
667 667 722 0 556 0 722 333 389 0 611 889 0 722 556 0 0 556 611 0 0 0 0 0 0 0 0 0 0 0 0 444
500 444 500 444 333 500 500 278 278 500 278 778 500 500 500 500 333 389 278 500 500 722 500
500 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 444
444]/BaseFont/TimesNewRomanPSMT/FirstChar 32/Encoding/WinAnsiEncoding/Type/Font>>□endobj□35
0 obj<</StemV 82/FontName/TimesNewRomanPSMT/FontStretch/Normal/FontWeight 400/Flags
34/Descent -216/FontBBox[-568 -307 2000 1007]/Ascent 891/FontFamily(Times New
Roman)/CapHeight 656/XHeight -546/Type/FontDescriptor/ItalicAngle 0>>□endobj□36 0 obj<</OPM
1/OP false/op false/Type/ExtGState/SA false/SM 0.02>>□endobj□37 0 obj<</Length
1914/Filter/FlateDecode>>stream
H‰”WK□âF□¾ó+êh¢Áã□`¤(R²›ÇDZíJá6“Cc□ÐÁã&Ýö□òëSÕm»□□□íaÇŒÙ©®ªïQÕ□_àûï□?}xú□□üðÃO□?À$,Ý$□
“Ç_ÿ ÏfòÓzò¸^G□Áz;‰Å(Š□X□ô□ë□ý€µ□8âç¿üJCœÐ3âÇ*,e”‡«□¬_'ÁGY□□`ºþkòózòó':éÑŸ□÷§_□Æç\Á[äy˜¸x÷B%
```

PDF Security Settings (3)



Brief Portable Document Format Archival (PDF/A) Tutorial

PDF/A – IS19005

- Subset of PDF for long-term preservation
- Nominal and full compliance
- Mandatory requirements

The PDF/A Standard

- ISO 19005 specifies how to use the Portable Document Format (PDF) 1.4 for long-term preservation of digital documents
- Applicable to documents containing character, raster, and vector data
- The standard does not address:
  - Processes for generating PDF/A files
  - Specific implementation details of rendering PDF/A files
  - Methods for storing PDF/A files

PDF/A Terminology

- PDF/A-1 refers to the format defined by Part 1 (ISO 19005-1) of the standard
- Part 2 (ISO 19005-2) will define PDF/A-2
- New Parts can be added to the PDF/A family of standards without obsoleting previous Parts

PDF/A

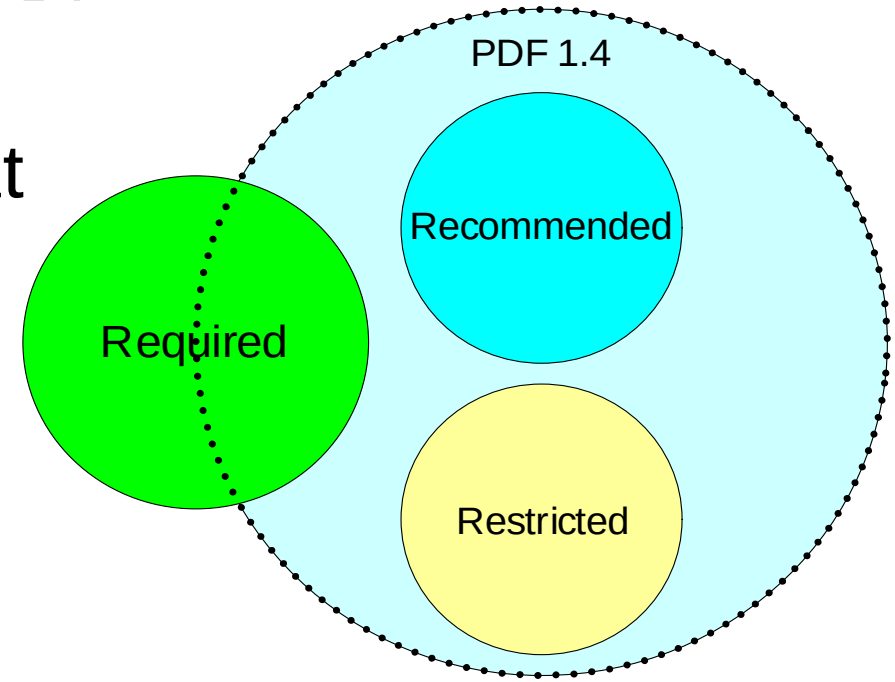
- Non-proprietary standard
  - Based on a proprietary, but open format
- Developed by inclusive set of stakeholders
- Subjected to rigorous technical review
- Minimal restrictions necessary to facilitate long-term preservation
- Not reliant on the existence of any particular reader

PDF/A

- PDF/A is intended to address three primary issues:
  - Define a file format that preserves the static visual appearance of digital documents over time
  - Provide a framework for recording metadata about digital documents
  - Provide a framework for defining the logical structure and semantic properties of digital documents

PDF/A Requirements

- Conformance to PDF 1.4
- With features that are
  - Required
  - Recommended
  - Restricted



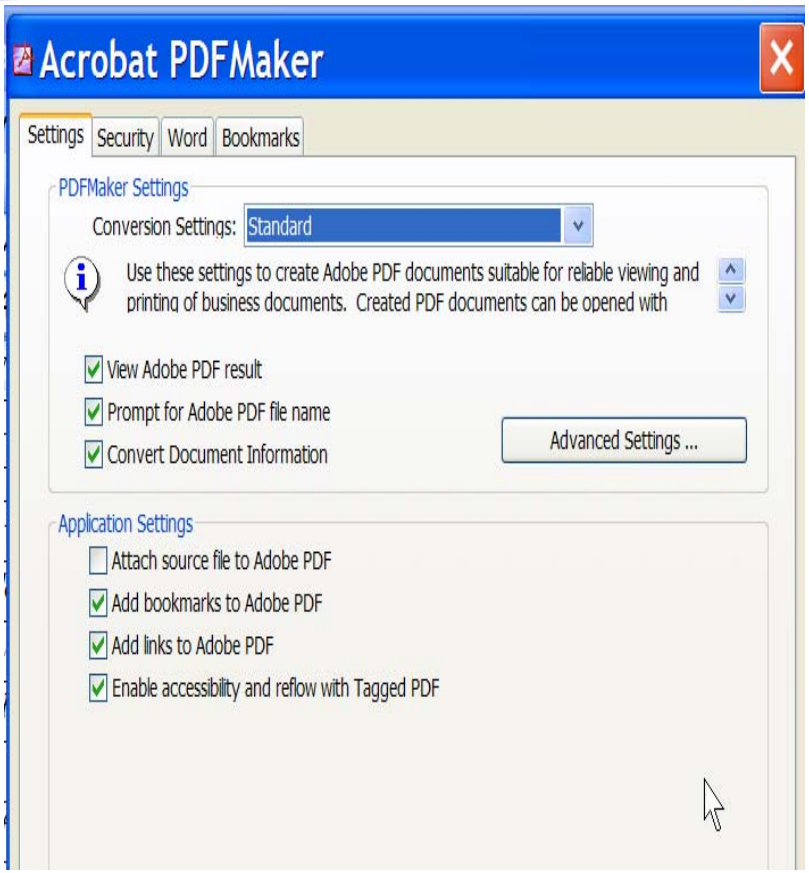
PDF/A Conformance

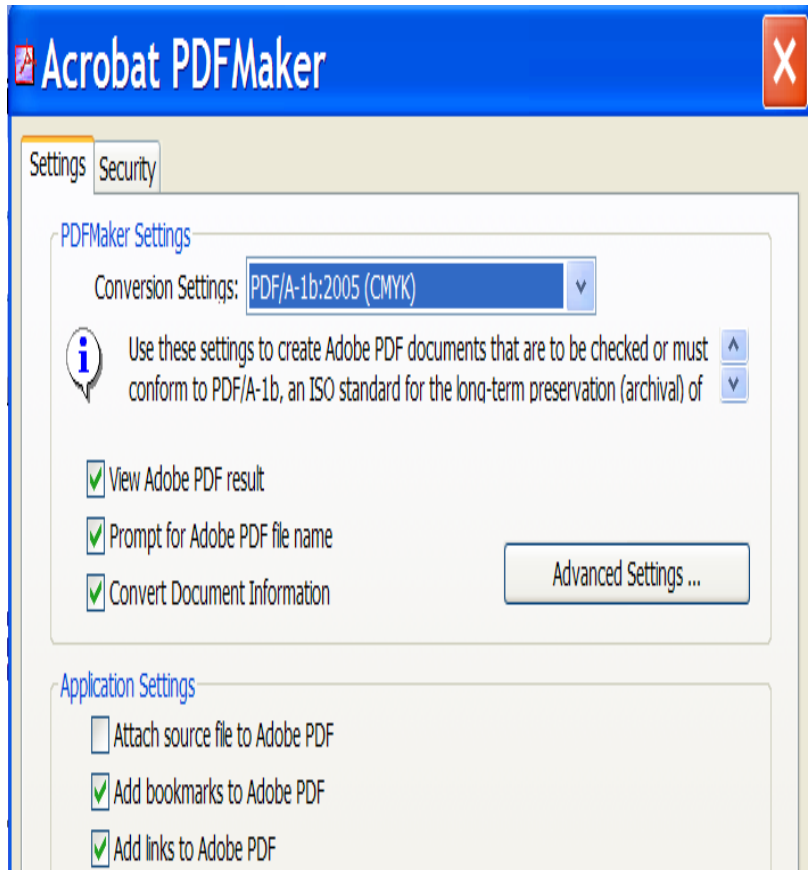
- Two conformance levels
  - PDF/A-1a
    - Compliance with all requirements of 19005-1
    - Including those regarding structural and semantic tagging
  - PDF/A-1b
    - Compliance with all requirements of 19005-1 minimally necessary to preserve the visual appearance of a PDF/A file

PDF/A Conformance Requirements

Features	Conformance	
	Minimal	Full
Required		
1. Device independent color	Yes	Yes
2. Embedded color spaces	Yes	Yes
3. Legally embeddable type fonts	Yes	Yes
4. Unicode	No	Yes
5. XMP	Yes	Yes
6. "Well formed" Tagged Documents	No	Yes
Prohibited		
7. Encryption	Yes	Yes
8. Password permissions	Yes	Yes
9. Lossy compression	Yes	Yes
10. LZW compression	Yes	Yes
11. All external dependencies	Yes	Yes
12. Sound and movie types	Yes	Yes
13. Executables such as JavaScript	Yes	Yes

How to Create PDF/A Records





PDF/A Conforming Software

- Apago PDF Appraiser
  http://www.apago.com

Original & Reproduced Vector Graphics

- Original vector graphics
  - Application native format
  - Visio
  - Corel Draw
- Reproduced vector graphics
  - Computer Graphics Metafile (CGM)
  - Scalable Vector Graphics (SVG)

Original & Reproduced Bit Map Images

- Original Bit Map Images
  - Digital cameras
  - Native format
- Reproduced Bit Map Images (scanned)
  - Tagged Image File Format (TIFF)
  - Joint Photographic Engineers Group
    - JPEG File Image Format (JFIF)
    - JPEF Still Image File Format (JSPIFF)
  - Portable Network Graphics (PNG)
  - Portable Document Format – image (PDF)

Original Text & Reproduced Text

Formats	Legal Admissibility Requirements					
	Accuracy of Capture	Integrity	Reliability	Human Readability	Accuracy of Reproduction	Trustworthiness
Original Text						
Plain ASCII	High	Low	High	High	NA	Medium
Word Processing	High	Low	High	High	NA	Medium
Marked-up	High	Low	High	High	NA	Medium
Reproduced Text						
Plain ASCII	NA	Low	High	High	Medium	Medium
RTF	NA	Low	High	High	High	Medium
HTML	NA	Low	High	High	High	Medium
XML	NA	Low	High	High	High	Medium
PDF	NA	High	High	High	High	High
PDF/A	NA	Medium	High	High	High	Medium

Original and Reproduced Vector Graphics

Formats	Legal Admissibility Requirements					
	Accuracy of Capture	Integrity	Reliability	Human Readability	Accuracy of Reproduction	Trustworthiness
Original Vector Graphics						
Native formats	High	Low	High	High	High	Medium
Reproduced Vector Graphics						
CGM	High	Low	High	High	High	Medium
SVG	High	Low	High	High	High	Medium

Original and Reproduced Bit Map Images

Formats	Legal Admissibility Requirements					
	Accuracy of Capture	Integrity	Reliability	Human Readability	Accuracy of Reproduction	Trustworthiness
Original Bit Map Images (digital photo)						
Native formats	High	Low	High	High	High	Medium
Reproduced Bit Map Images						
TIFF	High	High	High	High	High	High
JPEG-JFIF//SPIFF	High	High	High	High	High	High
PDF	High	High	High	High	High	High
PNG	High	High	High	High	High	High

Legal Admissibility Requirements and Electronic Records Formats

Formats	Legal Admissibility Requirements					
	Accuracy of Capture	Integrity	Reliability	Human Readability	Accuracy of Reproduction	Trustworthiness
Original Text						
Plain ASCII	High	Low	High	High	NA	Medium
Word Processing	High	Low	High	High	NA	Medium
Marked-up	High	Low	High	High	NA	Medium
Reproduced Text						
Plain ASCII	NA	Low	High	High	Medium	Medium
RTF	NA	Low	High	High	High	Medium
HTML	NA	Low	High	High	High	Medium
XML	NA	Low	High	High	High	Medium
PDF	NA	High	High	High	High	High
PDF/A	NA	Medium	High	High	High	Medium
Original Vector Graphics						
Native formats	High	Low	High	High	High	Medium
Reproduced Vector Graphics						
CGM	High	Low	High	High	High	Medium
SVG	High	Low	High	High	High	Medium
Original Bit Map Images (digital photo)						
Native formats	High	Low	High	High	High	Medium
Reproduced Bit Map Images						
TIFF	High	High	High	High	High	High
JPEG-JFIF//SPIFF	High	High	High	High	High	High
PDF	High	High	High	High	High	High
PNG	High	High	High	High	High	High

Take Away

- Few file formats inherently provide overall high support of legal admissibility
- Implications???

Part 4
File Formats and Digital Preservation

Desirable Properties of a Preservation Format

- Device independence
- No external dependencies
- Integrity
- Searchability
- Open standard
- Persistence
- Storage

Original and Reproduced Text

Format	Preservation Requirements						
	Technology Independent	External Dependency	Integrity	Searchability	Standard	Persistence	Storage
Original Text							
ASCII Text	High	None	Low	High	International	High	Low
Native word processing	Low	None	Low	High	Proprietary	Medium	Nominal
Market-up	High	None	Low	High	International	High	Nominal
Reproduced Text							
ASCII	High	None	Low	High	International	High	Low
RTF	Medium	Low	Low	High	Proprietary	Medium	Nominal
HTML	High	Low	Low	High	Internet	High	Nominal
XML	High	Low	Low	High	Internet	High	Nominal
PDF	High	Low	High	High	International	High	Low
PDF/A	High	None	Low	High	International	High	Low

Original and Reproduced Vector Graphics

Format	Preservation Requirements						
	Technology Independent	External Dependency	Integrity	Transferability	Standard	Persistence	Storage
Original Vector Graphics							
Native formats	Low	Low	Low	Low	Proprietary	Low	Unknown
Reproduced Vector Graphics							
CGM	High	Low	Low	High	International	Medium	Nominal
SVG	High	Low	Low	High	Internet	High	Nominal

Original and Reproduced Bit Map Images

Format	Preservation Requirements						
	Technology Independent	External Dependency	Integrity	Transferability	Standard	Persistence	Storage
Original Bit Map Images							
Native formats	Low	None	Low	Low	Proprietary	Low	Unknown
Reproduced Bit Map Img.							
TIFF	High	Low	High	High	International	High	High
JPEG-JFIF/SPIFF	High	Low	High	High	International	High	Nominal
PDF	High	Low	High	High	Proprietary	High	Nominal
PNG	High	Low	High	High	Proprietary	High	Nominal

Preservation Requirements and Electronic Records Formats

Format	Preservation Requirements						
	Technology Independent	External Dependency	Integrity	Transferability	Standard	Persistence	Storage
Original Text							
ASCII Text	High	None	Low	High	International	High	Low
Native word processing	Low	None	Low	Medium	Proprietary	Medium	Nominal
Market-up	High	None	Low	High	International	High	Nominal
Reproduced Text							
ASCII	High	None	Low	High	International	High	Low
RTF	Medium	Low	Low	Medium	Proprietary	Medium	Nominal
HTML	High	Low	Low	High	Internet	High	Nominal
XML	High	Low	Low	High	Internet	High	Nominal
PDF	High	Low	High	High	International	High	Low
PDF/A	High	None	Low	High	International	High	Low
Original Vector Graphics							
Native formats	Low	Low	Low	Low	Proprietary	Low	Unknown
Reproduced Vector Graphics							
CGM	High	Low	Low	High	International	Medium	Nominal
SVG	High	Low	Low	High	Internet	High	Nominal
Original Bit Map Images							
Native formats	Low	None	Low	Low	Proprietary	Low	Unknown
Reproduced Bit Map Img.							
TIFF	High	Low	High	High	International	High	High
JPEG-JFIF/SPIFF	High	Low	High	High	International	High	Nominal
PDF	High	Low	High	High	Proprietary	High	Nominal
PNG	High	Low	High	High	Proprietary	High	Nominal

Take Away

- File format selection should be driven by recordkeeping requirements
- Avoid proprietary single vendor products
- Use main stream technology products
- TIFF, XML, PDF, and PDF-A are good choices
- But what about integrity protection?

Inherent V. External Integrity Protection

- Inherent
  - Encryption
  - Password protected access to individual records (e.g., PDF)
- External
  - System Controls
  - Hash digest
  - Digital signature

System Controls

- Role based Security
- Records management architecture

Hash Digest Technology

- A fixed length (160 bits) digital fingerprint of string of 1s and 0s
- Computationally infeasible to extract a document from its digest
- Computationally infeasible to find an arbitrary message that has a particular digest
- Computationally infeasible for two messages to have the same digest
- Not reversible and sensitive to change of a single bit

Hash Digest Technology

- Hash algorithms
  - SHA- 1 (160 bits)
  - SHA-2 (224 bits)
- Pre/post activity comparison

Hash Digest Values

337.60 KB

AaAEAACoAQAKAGjhX8
4+VC1d3)NgDiPHvG+/R8
hKCAUCACOvWKATFOY
Iz3XS5gAAgI1wrAKO1ge
AAAAAAAAAAAAAAAA=

337.60 KB

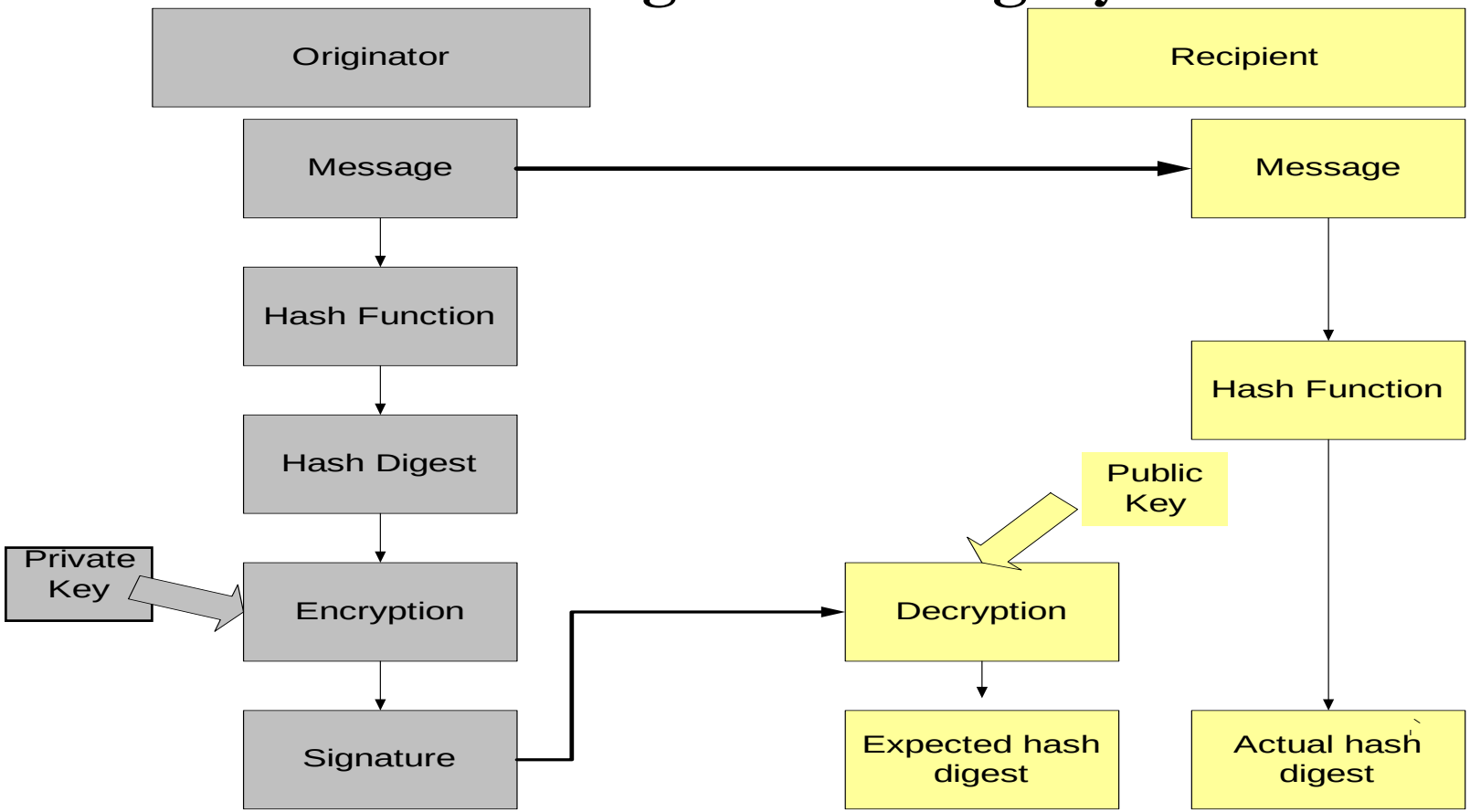
AaAEAACoAQAKAGy2YV
8gORjFeuf3yfnn7V)QMKB
CgKywNfTD+avB8UVEYK
AAAoUB2gKo1gEAALgAA
AAAAAAAAAA=

Digital Signature Technology

- Public Key Infrastructure (PKI)
  - Public key
  - Private Key
- Mathematical reciprocals
- Private key is pass word protected
- Obtain a digital certificate, publish public key
- Sign with personal private key
- Open with specific public of sender
- Individual or a server

Digital Signature Technology

Use of Hash Digest for Integrity Protection



Looking to the Future

- Electronic records preservation is never ending
- Volume of electronic records to be preserved is growing rapidly
- Do we need to preserve all electronic records as processibe records?
- John Palm, National Archives of Sweden, “The Digital Black Hole”

  http://www.tape-online.net/docs/Palm_Black_Hole.pdf

Estimated Five Year Cost to Preserve 25 TB of Records

Volume	Year 1	Year 2	Year 3	Year 4	Year 5
25 TB	45700	91500	137200	182900	228000
50 TB		45700	91500	137200	182900
75 TB			45700	91500	137200
100 TB				45700	91500
125 TB					45700
Total					675000

Feature Presentation

Part 2:

Lessons Learned: Hurricane Katrina, Dillard University

Dean McKinney, BMS CAT

Lessons Learned Hurricane Season 2005

Dillard University
New Orleans, Louisiana

MAJOR LOSS: Urban University Campus

e-Records Conference 2006

Historical Facts

- Over the years, this historic campus grew to 55 acres in the shadow of downtown New Orleans
- In recent history, the region has suffered through and sustained itself in the face of numerous hurricanes and serious weather events
- Never has the entire campus (and the city) been under water

Dillard University

Nick Harris
Assistant Vice President, Risk Manager

e-Records Conference 2006

BMS CAT

We thought we had planned for this…

- It had been discussed, even hyped for years…
- We contemplated the possibilities…
- We strived to anticipate the nightmare scenarios…
- We never believed that it would REALLY HAPPEN…
- THEN…

…IT REALLY HAPPENED!!!

- Worse than anyone had ever seriously anticipated…
- The ENTIRE campus was flooded for an extended period of time…
- No accessibility to the campus or off-campus housing
- Most inhabitants of the city were evacuated…
- No one really knew when New Orleans would “come home”…
- Our student body and all of our staff was gone too, with no projected return
- We never saw this played out on the Discovery Channel!!!!!!!!!!!!!

Here is just a taste…

No one really believed it would happen!

Dillard University

Nick Harris
Assistant Vice President, Risk Manager

e-Records Conference 2006

BMS CAT

BMS CAT

Disaster Planning

e-Records Conference 2006

What are the possibilities

- **Fires: You always anticipate fires and sometimes we plan for multiple facilities…but never the entire campus.**
- **Floods: You consider the possibilities of the obvious…pipe breaks, bad weather, storms and flash floods…but never the entire campus, much less the entire city! Not really!!!**
- **All the other stuff doesn't seem to matter when the unfathomable HAS HAPPENED!**

Dillard University

Nick Harris
Assistant Vice President, Risk Manager

e-Records Conference 2006

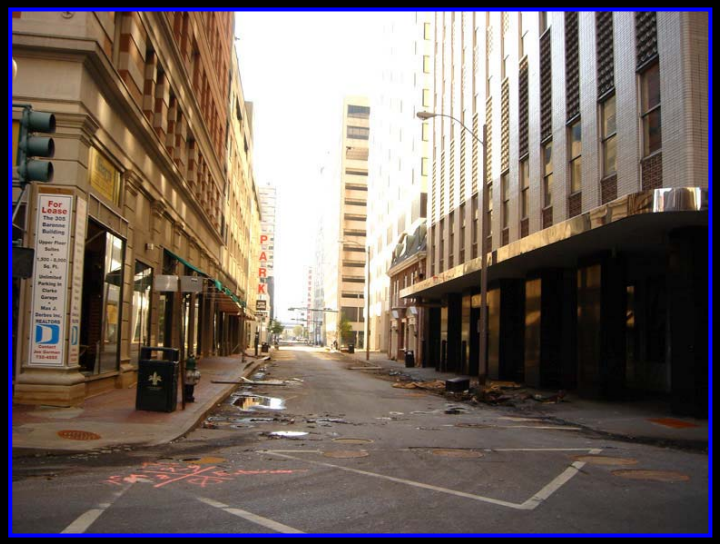
BMS CAT

Dillard University on the rebound

It Happened…water, water everywhere!!!

- Most of the city of New Orleans BUGGED OUT!
- About the time we thought we were spared the worst flooding, the levies broke (2 days later) and most of the city was underwater
- Floodwater was unabated in many areas of the city for WEEKS!
- Except for a few hearty souls in engineering (who were later evacuated) All of Dillard was gone.
- All 55 acres of Dillard University was flooded!!!
- The campus sat under water, to some degree, for 21 days!

New Orleans



It Happened…In the Dark!

- NO POWER anywhere for miles and NO IDEA when it would be restored!
- When power was restored (temporary generators) ON CAMPUS, it was the only electricity for miles. This made Dillard a beacon in the area. Not only one of hope but it also attracted undesirable elements to the ONLY light around.
- How do you get fuel to keep the generators running?
- To the amazement of all involved, Dillard remained under temporary power for MONTHS due to the complete loss of utilities in the area.

It Happened…In the Daylight!

- **Both local police and CAMPUS police were out of commission. Literally, there was no campus police – only hired security. Area constabularies were impacted for months and the National Guard was barely able to keep up with basic security and order.**
- **Looting was rampant in many areas, especially the NOLA-CBD. With communications, access and travel so limited, it was impossible to know what was happening at the campus for an extended time.**
- **To make the horror even worse, vandals set a flooded building on fire on campus… *yes, flooded buildings can burn!***

And that's the way it was…plus!

- Communications virtually non-existent
- No DRINKABLE water or sewer systems, other essential services indefinitely suspended
- No Food or any place to buy it or bring it in
- Kitchen facilities unusable – freezers and refrigerators has no power for weeks
- No suppliers and no deliveries
- Student body and staff displaced for months; some will never return to New Orleans or Dillard

Dillard University

Nick Harris
Assistant Vice President, Risk Manager

e-Records Conference 2006

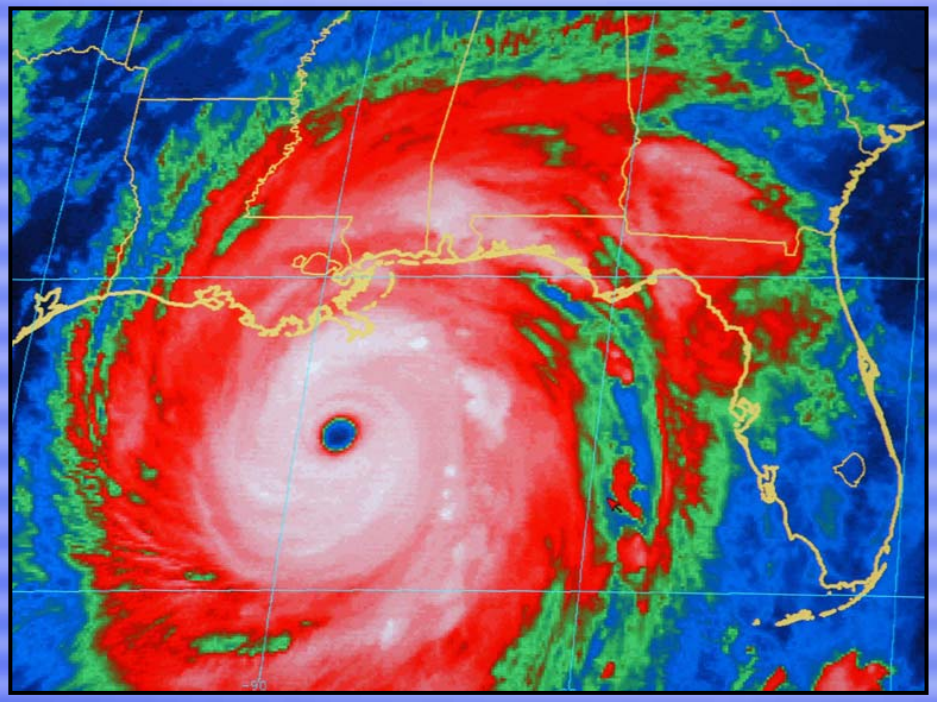
BMS CAT

And that's the way it was…plus!

- Campus deserted and offsite housing in even worse shape! Locked to try to stop the looters.
- Aside from most of the staff and student body being displaced, the very local people that you count on for labor and other services are GONE TOO! Many not to return, ever!
- Entire region of the state so severely impacted that all forms of relief are impeded from responding!
- At one point, the entire 504 area code was down!!

BMS CAT

So…what did we learn?



e-Records Conference 2006

Planning…

- You can never OVER-PLAN for potential destruction
- You can not fully anticipate what the level of damage actually can be
- When the disaster strikes. Move beyond PANIC-mode quickly and take CONTROL of the situation
- PUBLISH THE PLAN…everyone needs to know their role in the plan!

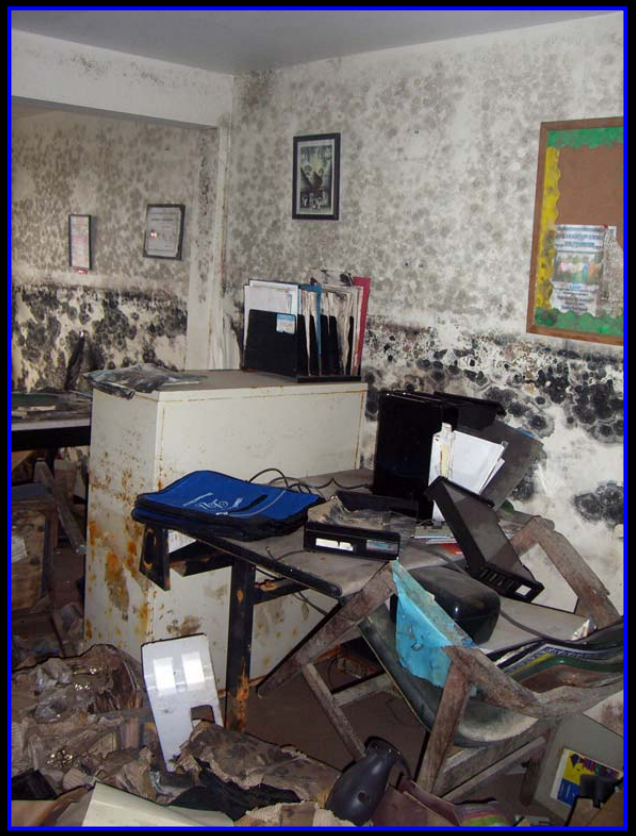
Response and Recovery…

- **How soon can you make it work?**
  - **What kind of time frame are you looking at for:**
    - **Response**
    - **Recovery (remember…these are different items)**
  - **What circumstances are affecting your timeframes?**
  - **Where can you go for help?**
  - **Solutions that are out of the box!**

e-Records Conference 2006

BMS CAT

Working on campus recovery



Security and Safety Issues…

- Police protection is non-existent as far as the campus force is concerned and spotty, at best, community-wide. SO WHO DO YOU CALL IN?
- Being faced with a “You against the cruel world” scenario, how do you prioritize your security needs and how are they taken care of?
- While basic law enforcement is of major concern, basic health and safety must be maintained:
  - First, so you are protecting your people, property and the recovery process and…
  - Second, so you don't get sued! Sad to say, but legal realities never stop. You can be sued.

Community Cooperation

- **Involve and utilize the various groups that routinely help you from outside the campus (URMIA, Local Safety Council, Fire and Police Departments, Chamber of Commerce, Churches and Civic Groups)**
- **Be sure that your local concerns and constituents do not feel pushed out but are a part of the solution.**
- **Influential Individuals that are concerned about your institution can be a huge help (A local city councilwoman was an alumnus of Dillard)**
- **CALL IN FAVORS…do not be shy…this is survival…help others when they need it and ask for it when you do! RELATIONSHIPS are the key!!!**

It is the GOLDEN RULE: Those that help you recover today are the ones you will be eager to help in the future.

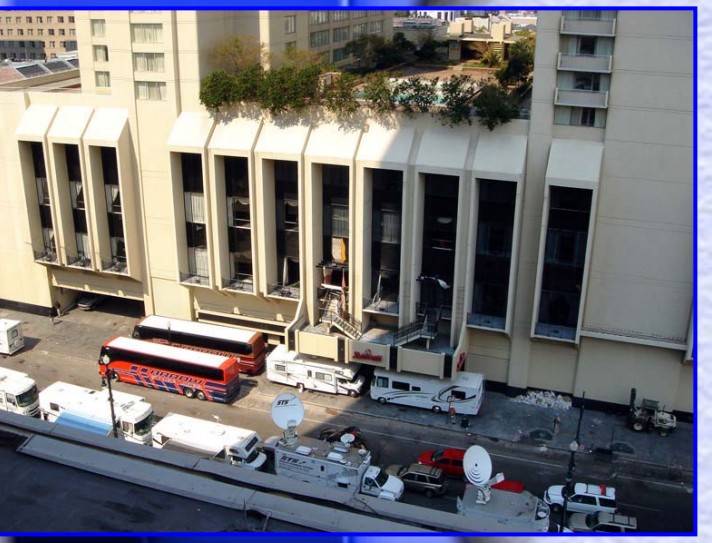
Recovery Implementation…

- **What are your GOALS?**
- **Where do you START?**
- **PRIORITIZATION is critical!**
- **Track your PROGRESS!**
- **Knowing where you have been tells you where you are going!**

e-Records Conference 2006

BMS CAT

Electrical Issues…



- No local power…no idea when it will come back on…this makes it very difficult to plan and respond.
- The power company will give you one story…the lineman in the field another…who to believe?
- Electricians are generally not hard to find…those that understand and know how to handle temporary power …ANOTHER MATTER ALTOGETHER!
- Adequate generators are expensive enough without your own power system being unable to handle the load and causing $680,000 in damage to a huge generator.

Water Issues…

- How do you clean a muddy mess with NO WATER?
- When you manage to obtain a water supply, how do you know it is safe? Who says? Are you sure?
- Irony of ironies….isn't it amazing that the 9 feet of water that ran thorough your campus made this mess and you can’t clean it up without water…now you can't get any!!!!
- How do you get rid of it?

e-Records Conference 2006

BMS CAT

Climate Control Issues…

- **Water damaged facilities must be stabilized eventually or they will mold again. It is one thing to demo/dry these buildings and quite another to stabilize the indoor environment for a healthy and safe building.**
- **Lack of electricity and water means NO chillers and limited or NO air-conditioning.**
- **In this case…Johnson Controls was a HUGE HELP even though their service contract had been terminated by the storm (Did you know that can happen?)**

Recovery of Property

- Student owned versus University owned property
- Faculty owned versus University owned property
- Mandatory evacuation imposed
- How do you allow people back on campus?
- Computers and other electronic equipment
- Artwork and special collections
- Communications systems

People’s stuff…

Dillard today and tomorrow…

- Dillard in on its way back and will be back stronger than ever.
- Presently, Dillard is occupying a large portion of the Hilton Hotel for student housing and classes as well as other locations in the area.
- The Historic Dillard campus is blooming again with enthusiasm, repairs and a lot of fresh paint.
- While having suffered a severe blow, they are counting their blessings and fully anticipate a wonderful future.

Three BIG lessons we all can learn from Dillard's experiences…

- **You can NEVER over anticipate or over plan for physical damage. While things are usually not as bad as they seem, sometimes they ARE WORSE!**
- **You can never underestimate the human impact of disasters, both in the initial effect of a catastrophe and the recovery phase. Sometimes people don't come back.**
- **Plan for the worst, ALWAYS. Train your people ALWAYS and when the unthinkable happens, do not be surprised when you come back stronger than ever!**

Your Turn…Q & A

BMS CAT

- **THANK YOU**

e-Records Conference 2006

BMS CAT

Contact Information

State and Local Records Management

Address: P.O. Box 12927
Austin, TX 78711-2927

Phone: (512) 452-9242

Fax: (512) 459-4967

Director and State Records Administrator
Michael Heskett x111
mheskett@tsl.state.tx.us

Program Planning and Research Specialist
Tim Nolan x126
tim.nolan@tsl.state.tx.us

Manager, Records Management Assistance
Mary Ann Bridges x135
mbridges@tsl.state.tx.us

Training Coordinator
Roy Bowden x154
rm_trng@tsl.state.tx.us

Publications Specialist
Kathryn Roberts x136
kathryn.roberts@tsl.state.tx.us

Government Information Analysts

Laura Finger x155
laura.finger@tsl.state.tx.us

Margaret Hermesmeyer x160
margaret.hermesmeyer@tsl.state.tx.us

Nanette Pfiester x137
nanette.pfiester@tsl.state.tx.us

Jed Rogers x157
jrogers@tsl.state.tx.us

Kim Scofield x158
kim.scofield@tsl.state.tx.us

Karon Teague x153
karon.teague@tsl.state.tx.us

State Agency Web Page
www.tsl.state.tx.us/slrm/state/index.html

Local Government Web Page
www.tsl.state.tx.us/slrm/local/index.html

DIR Contact Information

General Information

Executive Director
Larry A Olson
Chief Technology Officer - State of Texas
cto@state.tx.us

Main Office:
300 West 15th Street, Suite 1300
Austin, TX 78701
Phone: 512-475-4700
Toll Free: 800-348-9157
Fax: 512-475-4759
Email: dirinfo@dir.state.tx.us

Mailing Address:
P.O. Box 13564
Austin, TX 78711-3564

Public Information Officer
Thomas Johnson
Phone: 512-936-6592
Email: thomas.johnson@dir.state.tx.us

Open Records Requests/Approvals Office
Fax: 512-475-4759
E-mail: pio@dir.state.tx.us

Program Areas

DIR Store/IT Products and Services
www.dir.state.tx.us/servlet/dirStore
Customer Service: 800-464-1215

Contracts
Sherri Parks
Phone: 512-463-3580
E-mail: sherri.parks@dir.state.tx.us
www.dir.state.tx.us/store/faq/custfaq.htm

HUB Program
Bernadette Davis
Phone: 512-463-5712
Toll Free: 800-348-9157
E-mail: bernadette.davis@dir.state.tx.us

Telecommunications/TEX-AN
Phone: 512-463-3449
Fax: 512-463-3304
Complete Contact Listing:
www.dir.state.tx.us/tsd/contact.htm

Strategic Initiatives
Dustin Lanier
Phone: 512-463-7120
E-mail: dustin.lanier@dir.state.tx.us

Statewide Project Delivery
Program Lead: Rose Wheeler
Phone: 512-936-5900
E-mail: rose.wheeler@dir.state.tx.us
www.dir.state.tx.us/pubs/framework/index.htm

Enterprise Architecture Program
Program Lead: Barbara Nadalini
Phone: 512-463-5360
E-mail: barbara.nadalini@dir.state.tx.us
www.dir.state.tx.us/ea/index.htm

Policy & Research
Program Lead: David Dennis
Phone: 512-475-0577
E-mail: david.dennis@dir.state.tx.us

Security Office
www.dir.state.tx.us/tsd/security.htm