

# Digital Preservation

Tim Cornia

Utah Department of Technology Services

# Digital Preservation:

The process of ensuring that a digital object is accessible over the long term

The management of digital information over time

# Case Studies

## Local/Desktop Accounting

- ❖ Starts out as paper ledger (pre-1980's)
- ❖ Migrate to Lotus 1-2-3 on DOS (1980's-1990's)
- ❖ Moved to WordPerfect Quattro Pro in DOS (1990-1995)
- ❖ Moved to Excel 97 in Windows 95/98 (1995-1998)
- ❖ Moved to Excel 2000, 2004, 2008 (2000-2009)

# Case Studies

## Local/Desktop Accounting

What needs to be addressed to restore/migrate/access the data

- ❖ Do you have access to the data
  - Backups -Do you have backups? 5 1/4 floppy, 3.5 floppy
  - Is the data on a hard drive?
  
- ❖ Software
  - do you have the application software to access the data?
  - Do you have the Operating System to run the application on?
  - If using another program to open it, is your “other data” maintained (formulas, macros, etc) (emulation)
  
- ❖ Hardware
  - Do you have a PC to run the software (new PCs may not run the old OS or the application)
  - Can the hardware access your backup media (floppy)

# Case Studies

## Databases

- ❖ Databases typically resides on a server but may be on a workstation
- ❖ Application populating/accessing DB can exist on server, workstation, or both
- ❖ Databases can be upgraded without upgrading application
- ❖ Application can be upgraded without upgrading DB
- ❖ Upgrading either can cause unpredictable results

# Case Studies

## Databases

What needs to be addressed to restore the data

- ❖ Backups
  - Do you have backups? 9 track, 4 MM tape, 8 MM tape, other
  
- ❖ Database and applications (Digital obsolescence)
  - Do you have the DB software to access the data?
  - Do you have the application software to access the data?
  - Do you have Network Operating System to run the application on?
  - Do you have the workstation Operating System to run the application on?
  - Do you have the backup software to restore the data?
  
- ❖ Hardware (migration)
  - Do you have a PC to run the software (new PCs may not run the old OS or the application)
  - Do you have a server to load it on (with the correct Operating System)
  - Can the hardware access your backup media (tape size

# Case Studies

## Databases

What data do you need to restore?

- ❖ Field data – name, address, phone number, etc
- ❖ Transaction data
  - when was the file accessed last?
  - By whom?
  - When was the last update to the file?
  - By whom? Why?
- ❖ Security data
  - Names and passwords
  - When was the password reset
  - From what IP address was the app accessed?
  - From what IP address was the password reset?
  - What time was the name/password accessed
  - Who had rights to which parts of the DB and application?

# Digital Preservation

The question is not if you have backups or not. The question is, what do you need to restore?

- ❖ Digital preservation is not only about backups
  - Backups is one task of many in the digital preservation process
  - Restoration and accessibility of the data is another one of the parts of the digital preservation process
- ❖ The central piece tying backups and restorations together is policy, process, and procedure

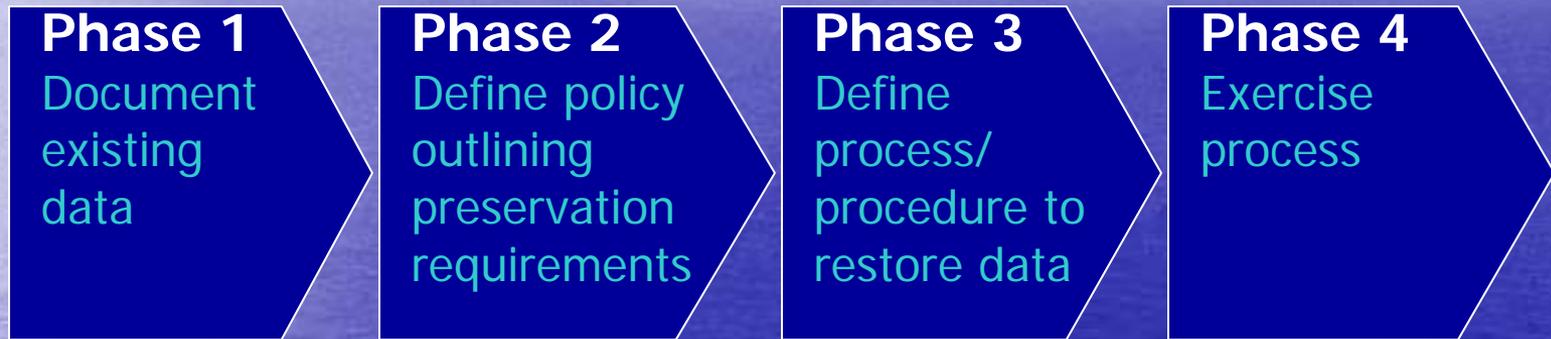
# Digital Preservation

Online Computer Library Center's (OCLC) four-point strategy for the long-term preservation of digital objects

- ❖ Assessing the risks for loss of content posed by technology variables such as commonly used proprietary file formats and software applications.
- ❖ Evaluating the digital content objects to determine what type and degree of format conversion or other preservation actions should be applied.
- ❖ Determining the appropriate metadata needed for each object type and how it is associated with the objects.
- ❖ Providing access to the content

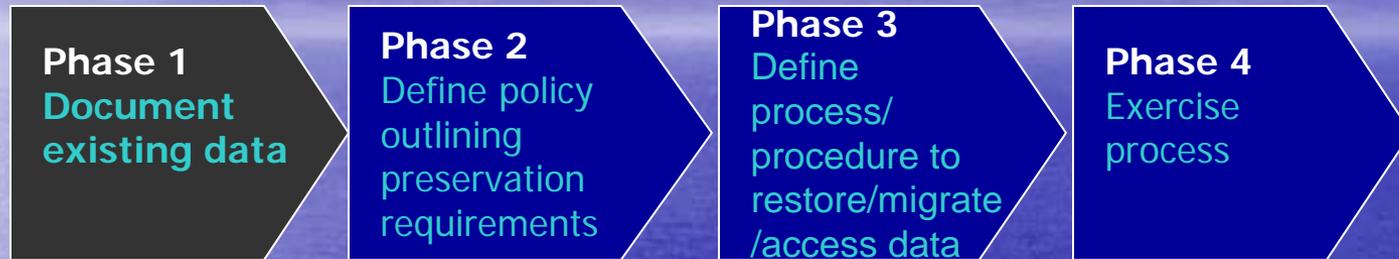
# Digital Preservation

## Mitigating digital preservation concerns process



# Digital Preservation

## Mitigating digital preservation concerns process



- ❖ Using a spreadsheet, document:
  - Each data type (customer DB, fiscal spreadsheet, video)
  - Format data was saved in
  - Application used to create data (Access 2000, Excel 2000, 2003, in house app)
  - OS on client or server
  - Special requirements to set up software (other software drivers needed to run software)

# Digital Preservation

## Mitigating digital preservation concerns process



- ❖ The policy should address the following:
  - How far back do you need to be able to restore data? Reference federal requirements and business documents here to justify the “why”
  - What format should the data be saved in?
  - What format does the data need to be restored to?
  - What data needs to be restored? (data versus photos of data, imaging)
  - Timelines to restore data

# Digital Preservation

## Mitigating digital preservation concerns process



- ❖ The policy should address the following:
  - How far back do you need to be able to restore data? Reference federal requirements and business documents here to justify the “why”
  - What format should the data be saved in?
  - What format does the data need to be restored to?
  - What data needs to be restored? (data versus photos of data, imaging)
  - Timelines to restore data

# Digital Preservation

## Mitigating digital preservation concerns process



- ❖ Some general principles should be followed when creating policy in this area:
  - a policy needs to convey the very philosophy of an organization concerning digital preservation; it should induce a common understanding of the objectives, of whether each collection item should be preserved with maximum effort possibly applying multiple preservation paths, or whether a certain pragmatism should be pursued;
  - a digital policy should facilitate the sustainability of an institution's present and future digital holdings;
  - a digital preservation policy has to demonstrate its benefits, its effectiveness;
  - a digital policy should be connected and integrated with a risk assessment document;

# Digital Preservation

## Mitigating digital preservation concerns process



- ❖ Some general principles should be followed when creating policy in this area (continued):
  - every policy should be practicable, not definitive, capable of being put into practice by institutions with varying resources and needs, and, especially, flexible to adapt itself to changing administrative and technological circumstances;
  - any policy should be characterized by clarity, adequacy, transparency, efficiency, effectiveness and logical organization of contents;
  - a digital preservation policy should be written in a simple and suitable language, without redundancies and, at the same time, without lowering the level of quality contained in its contents; (ERPANET)

# Digital Preservation

## Mitigating digital preservation concerns process



- ❖ The process/procedure should address the following:
  - Hardware, software, OS needed to restore data including media hardware, backup software, etc
  - Who is responsible for what and by when?
  - What form does the data need to be restored to (data versus photos of data)
- ❖ Business and IT should work together to make sure expectations are clearly defined and understood so that the processes/procedures meet the needs of the business

# Digital Preservation

## Mitigating digital preservation concerns process



- ❖ The testing of the process should address the following:
  - Verifying the hardware, software, and documentation to restore/access the data functions as planned
  - Verifying the process/procedure covers all the possible plausible contingencies
  - Verifying the data is restored/migrated in the form expected by the business (data, verses photos of data)
- ❖ Business and IT should work together to make sure expectations are clearly defined and understood so that the processes/procedures meet the needs of the business

# Digital Preservation Technologies



# Digital Preservation Technologies



- ❖ Use standard formats that will be supported long term (Tiff)
- ❖ There are different proprietary digital camera formats. Stay with open standard formats. If saving in RAW, make sure it is not proprietary
- ❖ Realize that scanning data is not the same as having the data. A Scanned spreadsheet is a photo of the result of the data, not the data itself

# Digital Preservation Technologies



- ❖ Hard disk storage keeps falling in price.
- ❖ Data can be stored on inexpensive SATA drives if access speed is not an issue
- ❖ Every time your backup systems are replaced, the question that needs to be asked is "are we migrating our backed up data to the new backup system"
- ❖ Work closely with IT to make sure your business needs are articulated (policy document)

# Digital Preservation

❖ Questions?