Digital Preservation Cross Discipline Survey

Stacy Kowalczyk
SLIS Ph.D. Conference
9/24/2005

Digital Libraries and Preservation

Since 1994, libraries have been developing a body of research and practice to preserve the materials that are either digitized for better access or "born digital"

- ❑ Mellon Foundation
- ❑ Digital Library Federation
- ❑ Library of Congress
- ❑ National Science Foundation

Digital Preservation

Preservation Goals

- Keep the bits safe
- Keep the files useable
- Keep the integrity of the object
- Keep the context of the object

Requires an active, systematic program
(Waters & Garrett 1996)

Keep the Bits Safe

- File maintenance
  - Regular error checking
  - Inventory checks
  - Checksums
- Regular backups
- Media migration
  - Degradation
  - Obsolescence
- Disaster Recovery Planning
  - Off-site backups
  - Transaction logging

40 years of best practice just needs to be implemented.

Keep the Files Useable

This is much harder…

- Depends on the complexity and transparency of file properties
    - File formats
    - Compression algorithms
- Requires significant knowledge and data
- Requires on-going management

…and is the current focus of major research efforts.

Digital Preservation Dichotomy

- In research and practice journals
    - The need is obvious
    - The problem is real and imminent

But…

- In the popular press
    - Anything digital is better than anything physical
    - Digital is the preservation standard

PARADE

Columns | Specials | Cover Story | E-Mail PARADE | Help | Archive | WebLinks

PARADE**Archive**

Archive Search

Articles are archived one week after publication.
▶ About the Archive

Back to April 25, 2004 index | Back to archive home

Preserve Your Family Memories
By Robert Moritz
Published: April 25, 2004

Isn't it amazing how many memories you can fit in a shoebox? Snapshot after snapshot of your favorite family moments—soccer games, weddings, doggie Halloween contests—all crammed together, gathering dust and fading with time. Wouldn't it be great if you could easily restore them to pristine condition, organize them by category, create great-looking photo albums and then easily share it all with friends and family? Well, thanks to the power of your PC, you can.

From Paper To Pixel s
The first thing you need to do is convert your old prints into digital photos. To do this, you'll need a scanner. The good news is that high-quality scanners can be had these days for less than a hundred bucks and set up in less than 15 minutes. Look for a model that includes a transparency adapter for scanning from slides and negatives: It's always better to scan directly from the original negative when you can. Epson puts out a first-rate machine for newbies (see below).

Search No More
To get the most out of your scanner—and to keep your sanity as your hard drive fills up with images—I recommend using photo-management software. These programs wear a variety of hats, including working with the scanner to capture new images, scouring your hard drive for existing pictures taken with a digital camera, touching up image quality and organizing your photo collection. The most complete management program, in my opinion, is Adobe Photoshop Album 2.0 ($50). The beauty of this program is its ability to cross-categorize so that individual pictures can be marked with multiple "tags"—such as "Summer 1986," "Grandpa," "Birthday Parties"—allowing for quick searching and sorting of related shots. It also has a nice selection of "Auto Fix" buttons for doing everything from banishing red eye to brightening dark snapshots and cropping out ex-boyfriends. Another comprehensive organizing program is the Magix Photo Cleaning Lab (see below) featuring Photo Clinic 3.0 for removing scratches, tears and dust from scanned images as well as creating panoramic montages.

Share The Memories
What's the point of restoring and organizing family memories if you don't spread the love? Sharing digital images is simple these days, thanks to a bumper crop of software and Web sites that do the heavy lifting for you. A great place to start is

Most Popular Stories
- Heart Galleries
- Ask Marilyn Daily Diet
- Do You Know *All* Your Deductions?
- Cinnamon And Diabetes
- Home Defibrillators
- The Best Medical Help Online
- When Your Insurance Won't Pay
- Materials Exchanges
- Are You Fluent in 'Feline'?
- He Turns Boys Into Men

A Personal Encounter

The Research Question

Is there really a Digital Preservation dichotomy between academics and the regular world?

- What is the level of awareness of digital preservation among computer savvy people?
- What is the level of knowledge of digital formats in these same computer savvy people?

Operationalizing the problem

- Created a survey with 26 questions
  - 11 on attitudes and awareness
  - 15 on knowledge of specific digital formats
- Surveyed IU students in computer – related classes
  - Self-identified demographics
    - Area of Study
    - Level of study
    - Gender

Participants

- 229 participants
- With 220 valid responses
- From 15 classes
  - 5 SLIS classes
  - 2 Informatics classes
  - 5 CS classes
  - 3 Business classes
- One email list (for more CS and Informatics Grad students)
- One visit to an Informatics computer lab

Population Breakdowns

	Business [N=50] 23%		CS [N=40] 18%		Informatics [N=51] 23%		IS [N=33] 15%	LS [N=31] 14%	Other [N=12] 5%	
	Grad	Undergrad	Grad	Undergrad	Grad	Undergrad	Grad	Grad	Grad	Undergrad
Male	19	16	7	18	3	28	13	9	1	5
Female	4	1	6	1	5	4	10	20	1	4
N/S	6	4	1	7	0	11	10	5	0	1
Totals [N=220]	29	21	14	26	8	43	30	34	2	10

Data

For each of the statements presented below, please indicate your opinion by placing an X in the appropriate column.

- Awareness questions were coded for an optimism factor
    - Strongly agree = 2
    - Agree =1
    - Disagree = -1
    - Strongly disagree = -2
    - No opinion = 0

Awareness Section

Category	Survey Questions
Longevity of digital information	1. Information in digital format will last longer than information on paper. 2. Digital photographs will last longer than film. 4. Digital image formats are stable and will last a long time. 10. Library and Museum materials (images, text, etc) are safer in electronic format.
Confidence in software	6. If the format that my digital camera uses goes out of date, I am sure that I will be able to upgrade my photos to a new format. 7. If I am audited by the IRS in 2010, I will be able to use my 2005 TurboTax software and data to recreate my tax filings. 8. Businesses migrate data from one system to another and from one record format to another with no adverse affect on their customers.
Availability of access	3. Academic journals that are published electronically today will still be available online in 10 years. 9. Information published on Websites will be accessible to researchers in 10 years.
Personal practice	5. I know the technical format that my digital camera uses. 11. I back up my computer files regularly.

Data

If you had to store your work in a digital format and could not touch it for 10 years, which formats would you choose? For each of the digital formats listed below, please indicate with an X in the appropriate column how safe you think that the format is for long term storage.

- Knowledge questions were coded for a safety scale
    - Very safe = 2
    - Safe =1
    - Risky = -1
    - Very Risky = -2
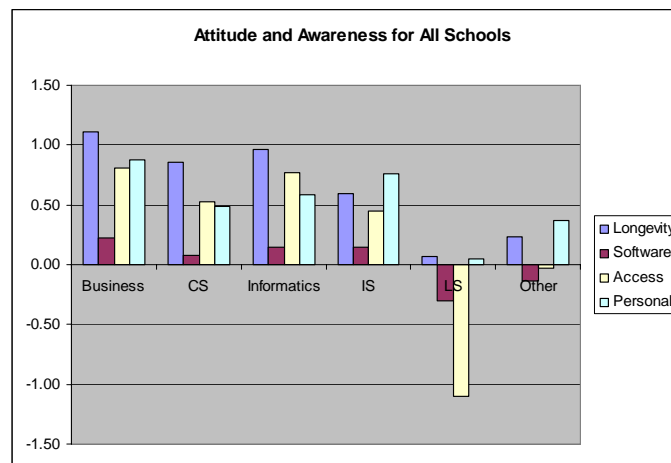   - No opinion = 0

Judging Knowledge

- Determined three categories based on best practice
    - National Archives and Records Administration (NARA)
    - The National Information Standards Organization's (NISO)
    - The Library of Congress
      (NARA, 2004; NISO, 2004; Arms & Fleischhauer, 2005)

- Seven sustainability factors of Library of Congress
    - Disclosure
    - Adoption
    - Transparency
    - Self-documentation
    - External dependencies
    - Impact of patents
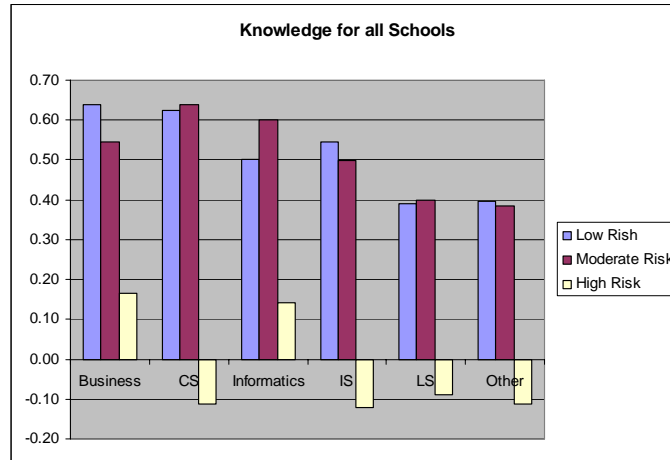   - Technical protection mechanisms   (Arms & Fleischhauer, 2005)

Format Categories

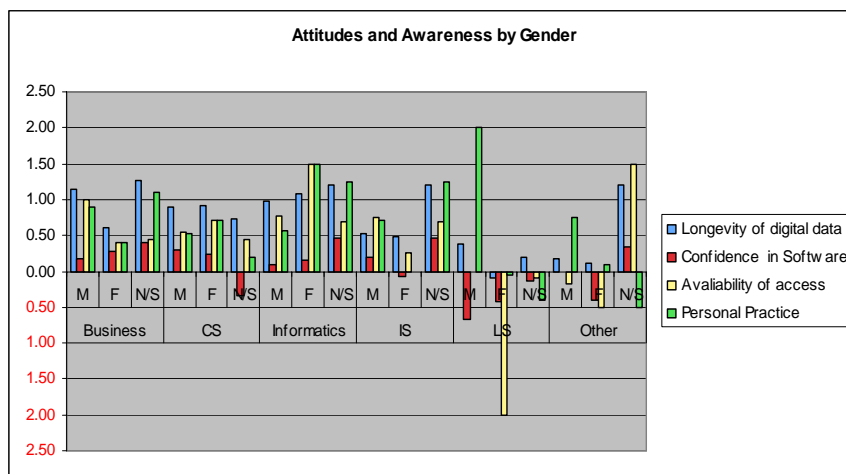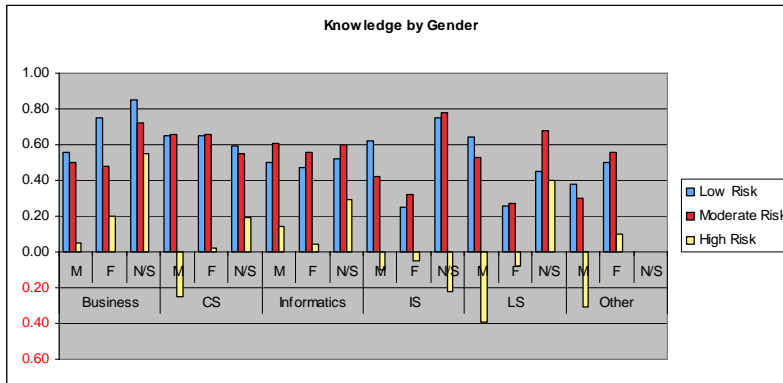| Category	Format
Low risk formats *Loss-less and transparent*	TIFF, JPG2000, ASCII, XML
Moderate risk formats *Lossy, semi-transparent, deprecated, or not widely adopted*	JPG, PNG, RTF, HTML, SGML
High risk formats *Lossy and opaque*	GIF, PhotoCD, MSWord, WordPerfect, ClarisWorks, MS Excel

Overall Attitudes and Awareness
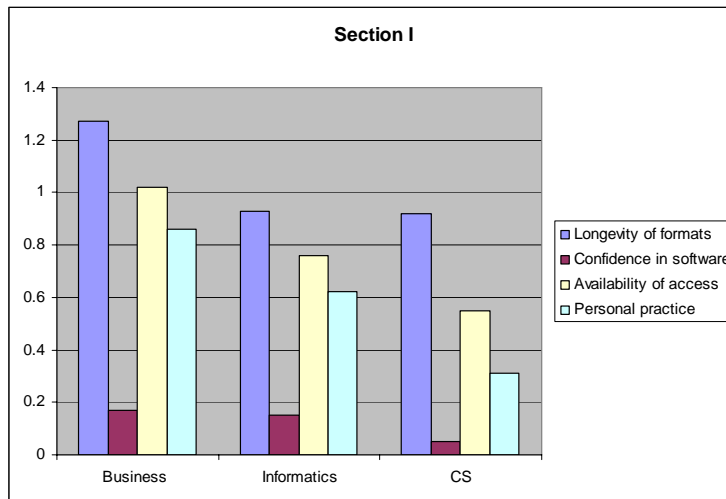
Overall Student Knowledge

Knowledge for all Schools

Legend: Low Rish, Moderate Risk, High Risk

Categories: Business, CS, Informatics, IS, LS, Other

Awareness Summary by Gender

Attitudes and Awareness by Gender

Legend: Longevity of digital data, Confidence in Softw are, Avaliability of access, Personal Practice

Categories (M, F, N/S): Business, CS, Informatics, IS, LS, Other

Knowledge Summary by Gender

Knowledge by Gender

Legend:
- Low Risk (blue)
- Moderate Risk (red)
- High Risk (yellow)

Categories: Business, CS, Informatics, IS, LS, Other — each with M, F, N/S

Y-axis: 1.00, 0.80, 0.60, 0.40, 0.20, 0.00, 0.20, 0.40, 0.60

Undergraduate Awareness Summary

Section I

Legend:
- Longevity of formats
- Confidence in software
- Availability of access
- Personal practice

Categories: Business, Informatics, CS

Y-axis: 0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4

Undergraduate Knowledge Summary

Section II
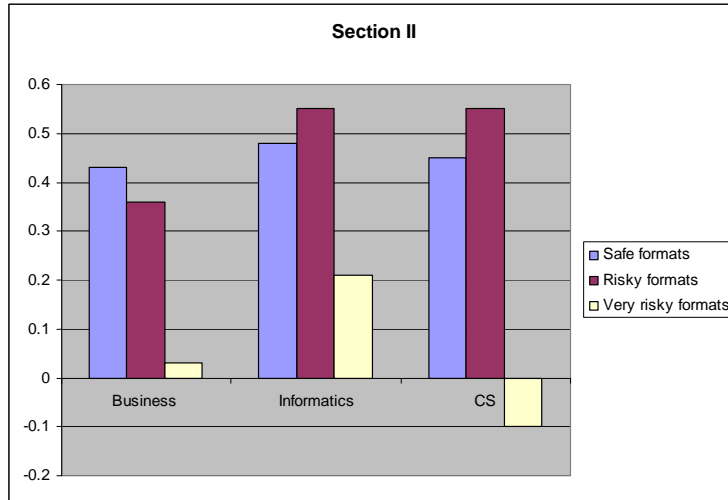
Graduate Students Awareness

Attitudes and Awareness Summary for Graduate Students
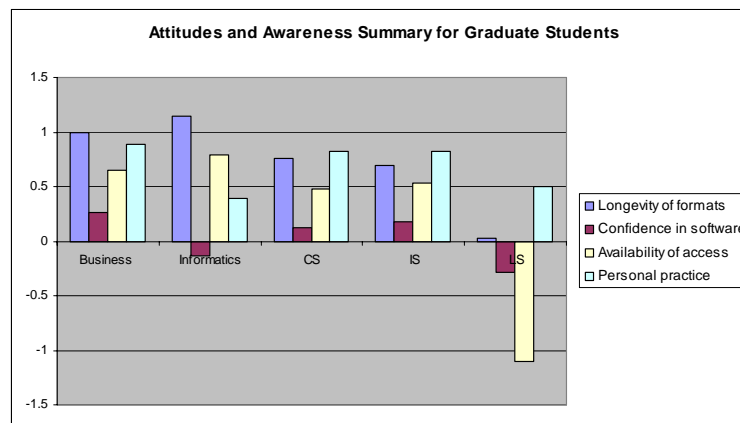
Graduate Students Knowledge

Knowldege Summary for Graduate Students

Legend:
- Safe Formats
- Risky Formats
- Very Risky Formats

Categories: Business, Informatics, CS, IS, LS
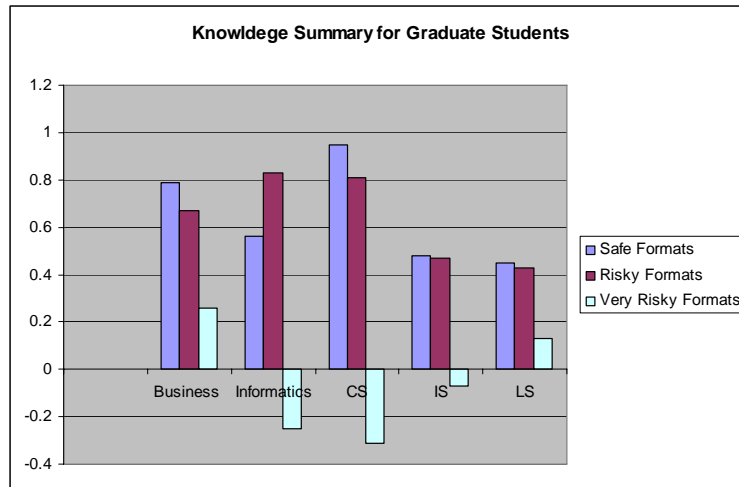
Results Summary

- In general,
  - Undergraduates are more optimistic than graduate students
  - Business majors, both undergrad and grad are the most optimistic
  - Informatics students are nearly as optimistic as business students

- Most students think that
  - Formats will persist
  - Access will persist
  - Business will not migrate data well
  - Software will not be able to read old data

- Library Science students are significantly more pessimistic than all other students

Discussion

❑ Optimism continuum that seems to support a digital preservation dichotomy

❑ Education level seems to increase format knowledge and lower optimism levels

❑ Paradoxically, low trust in businesses and software do not equate to low trust in proprietary formats
  ■ Microsoft formats rated as safe

❑ Format familiarity equates to format safety
  ■ HTML rated safer than XML
  ■ JPEG and GIFF formats rated safer than TIFF

References

Arms, C. R. & Fleischhauer, C. (2005). *Digital formats for library of congress collections.* Retrieved on March 28, 2005 from http://www.digitalpreservation.gov/formats/intro/intro.shtml

Babbie, Earl. 2001. *The practice of social research, 10th Edition.* Belmont, CA: Wadsworth.

Hart, P., Liu, Z. (2003). Trust in the preservation of digital information. *Communications of the ACM, 46*(6), 93-97.

NARA. (2004). *Technical guidelines for digitizing archival materials for electronic access: creation of production master files - raster images.* Retrieved on March 28, 2005 from http://www.archives.gov/research_room/arc/arc_info/guidelines_for_digitizing_archival_materials.html.

Marcum, D. B. (2001). CLIR forms planning group for preservation survey. *CLIR Issues, 20.* Retrieved on April 19, 2005 from http://www.clir.org/pubs/issues/issues20.html.

Moritz, R. (2004) Preserve your family memories. *Parade Magazine,* April 25, 2004. Retrieved on March 2, 2005 from http://archive.parade.com/2004/0425/0425_family_living.html

NISO Framework Advisory Group. (2004). *A framework of guidance for building good digital collections. 2nd edition.* Bethesda, MD: National Information Standards Organization. Retrieved on March 20, 2005 from http://www.niso.org/framework/framework2.html

SPSS for Windows, Rel. 12.0.2. 2004. Chicago: SPSS Inc.

Waters, D., &Garrett, J. Eds. (1996). *Preserving digital information: Report of the task force on archiving of digital information.* Washington, D.C. and Mountain View, CA: The Commission on Preservation and Access and the Research Libraries Group. Retrieved March 4, 2005 from http://www.rlg.org/ArchTF/

Williams, R. (2004). Trusting "corporate memory". *KM World, 13*(9).

Williams, R. F. (2003). *Electronic records management: A call to action.* Retrieved on April 21, 2005 from http://www.merresource.com/whitepapers/pdf/survey.pdf.