

Digital Preservation Cross Discipline Survey

From newspapers to photography magazines, the popular press is touting digital technology as the best method for individuals to preserve their photographs while computer scientists, information scientists, librarians, and archivists are writing article after article in academic journals lamenting the impending loss of knowledge because the lack of a comprehensive digital preservation strategy.

Since 1994, libraries and archives have been developing a body of research and practice to preserve the materials that are either digitized for better access or “born digital.” Since 1996 when the seminal paper on digital preservation was published, the fundamental discussion was framed by the preservation goals: to keep the bits safe, to keep the files useable, to keep the integrity of the object, and to keep the context of the object (Waters & Garrett 1996). While the Sunday newspaper magazine section states under the headline of “Preserve Your Family Memories:”

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 (Moritz 2004).

I have experienced this great divergence between the popular media and the general public’s view of technology and the information scientists’, librarians’, and archivists’. This past Thanksgiving, I was rummaging through a box of photos and discovered several daguerreotype portraits dated 1852. I remarked to my brothers, one an electrical engineer and the other a director of information technology at a major publishing house, that future generations will not just happen upon old photographs that have been preserved by benign neglect; our digital photos will take

regular action to preserve. Both of them looked at me with great puzzlement; the director of IT told me that he had a good hard drive and backed up his computer regularly – what’s the problem?

My brothers think that “keeping the bits safe” is good enough. If multiple copies exist in multiple locations, if the storage media is monitored for data degradation or technological obsolescence, if data center best practice that has been developed over the past 40 years is followed, then the data is “preserved.”

Unfortunately, “keeping the bits safe” is only the first, and the easiest, step in digital preservation. The longevity of digital information depends on more than good backups. For individuals with a computer full of digital files, the second of the preservation goals is just as important. Without understanding the relative risks of data formats, not only will individuals lose their financial, intellectual, and family histories but information scientists, librarians, and archivists will not be able to develop serious technology solutions or public policy if the problem is not recognized as real and imminent. We need to do more than collect anecdotal family stories; we need to know what people know about digital information and what their attitudes about digital preservation are.

Background

What exactly is known about peoples’ attitudes towards digital information?

Libraries have been studied. The Council on Library and Information Resources conducted a study of library and archives to identify preservation projects (traditional as well as digital) and digitizing projects, as well as to determine the preservation budget and staffing resources (Marcum, 2001).

Businesses have been studied. Biannually between 1999 and 2003, the Cohasset Group, an international business consulting company, has surveyed records management professionals via two professional organizations. The most recent survey’s results showed that a great majority of

records managers are not confident in their ability to reproduce legally sufficient documents from their records. (Williams, 2002 and 2003). Hart and Lui conducted a survey of 110 computer power users to determine their level of trust in electronic records. Of the respondents, 86% state that they would keep paper copies of important digital documents. The authors have determined that there are 5 reasons that people do not trust their electronic documents: *inaccessibility* – people want to be able to get to their data without barriers of electrical power, hardware reliability or reading devices; *lack of tangibility* – people want a solid thing, not invisible bits on a opaque disk; *fluidity* – people feel that the electronic copy can be easily altered while the paper copy is fixed in time and space; *short preservation period* – people understand that the physical media of digital storage is short lived; *privacy and security* – people think that electronic information can be more easily compromised than paper (2003).

But none of these studies helps to determine if there really is 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?

Methods and Design

To help answer these questions, a survey was developed to gather information about the attitudes and knowledge of computer savvy students at IU Bloomington. Students were recruited through computer-related classes which were selected from the Kelly School of Business, the School of Library and Information Science, and the School of Informatics (which includes the computer science department). Courses selected ranged from introductory classes in all programs to the most advanced. (See Appendix B for the complete list of courses surveyed.) To collect sufficient Computer Science and Informatics graduate students, two additional methods of

recruiting students were used – the WIC-ALL email list and the Graduate Informatics Computer Lab which resulted in an additional 8 surveys. Nine of surveys were not completed. Those surveys were not counted.

For this study, the students were allowed to self-identity an area of study. Business, Computer Science (CS), Informatics, Information Science (IS), and Library Science (LS). Many students identified multiple areas of study. These surveys were coded by the area of study closest to the course in which the survey was given. As an example: a student self-identified both business and computer science; if the student took the survey in a business class, s/he was coded as a business student; if the student took the survey in a computer science class, s/he was coded as a CS student.

Survey Population

	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	Under grad	Grad	Under grad	Grad	Under grad	Grad	Grad	Grad	Under grad
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

The students, who were anonymous but did provide categorizing information – school, education level, and gender, were given a survey with 26 questions that could be completed in less than 10 minutes at the end of class. The survey had two sections with questions that were designed to gather information about both attitudes and knowledge. Both sections of the survey used a semantic differential scale. Section One, the attitude section, used an agreement scale – highly agree, agree, disagree, strongly disagree. Section Two, the format preservation quality section, used a risk scale – very safe, safe, risky, very risky. The data was coded from positive to negative – +2

to -2. The “no opinion” option or no answer was coded as zero. The survey questions and summary results can be found at Appendix A.

Section I had 11 questions about student attitudes in 4 aspects of digital preservation: longevity of digital information, confidence in software, availability of access, and personal practice. The first category, longevity of digital information, asks questions about how permanent or persistent the students think digital information will be. The second category, confidence in software, asks questions about the ability of software over time to perform over time, the availability of software to convert data and business’ ability to provide seamless conversion services. The third category asks questions about the long-term accessibility of the intellectual of digital information. The fourth category asks about the students’ own personal practice of good digital preservation practice.

Question Categories

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.

Section II had 15 questions asking the students to rank digital data formats’ preservation quality. The formats have been grouped according to the current best practice as described in the

National Archives and Records Administration (NARA), the National Information Standards Organization's (NISO), and the Library of Congress' (LOC) digitizing recommendations (NARA, 2004; NISO, 2004; Arms & Fleischhauer, 2005). How do these organizations determine what is a safe or risky file format? The Library of Congress has developed a theoretical framework for assessing formats using seven sustainability factors – disclosure, adoption, transparency, self-documentation, external dependencies, impact of patents, technical protection mechanisms – which influence the probability of preserving the intellectual content of the digital objects (Arms & Fleischhauer, 2005).

Format Categories

Category	Format
Safe formats <i>Loss-less and transparent</i>	TIFF, JPG2000, ASCII, XML
Risky formats <i>Loss, semi-transparent, or not widely adopted</i>	JPG, PNG, RTF, HTML, SGML
Very risky formats <i>Lossy and opaque</i>	GIF, PhotoCD, MSWord, WordPerfect, ClarisWorks, MS Excel

Results

The data was analyzed to determine two scores – one for attitudes and one format risk. The attitude scores are based on an optimism scale from +2 (most optimistic) to -2 (least optimistic). The format score is a safety scale from +2 (most safe format) to -2 (least safe format). The scores were both calculated by summing the scores and dividing by the number of entries. This average includes the “no opinion” entries.

Survey Summary Results

Section I Question Categories	Total [N=220]	Female [N=56]	Male [N=119]	Not Specified [N=45]
Longevity of digital information	0.71	0.34	0.81	0.92
Confidence in software	0.07	-0.18	0.14	0.19
Availability of access	0.53	0.19	0.69	0.52
Personal practice	0.57	0.25	0.68	0.66

Section II Format Category	Best Practice	Totals [N=220]	Female [N=56]	Male [N=119]	Not Specified [N=45]
Safe formats	1.5	0.54	0.40	0.57	0.64
Risky formats	- 0.5	0.53	0.42	0.55	0.65
Very risky formats	-2	0.01	0.01	-0.06	0.22

Gender identification seems to be a significant issue with the students. Nearly 25% of the participants (45 out of 220) did not specify gender while only .05% (1 out of 220) did not identify area of study. When this phenomenon was first noticed, the researcher hypothesized that the females did not want to be identified, so the surveys were scanned for gender as they were being handed in. While certainly not scientific, the researcher noticed that males as well as females were not self-identifying gender. When leaving a classroom with only males students, nearly 50% of the surveys for that class did not have gender marked. Gender then becomes a difficult variable to use as a discriminator in the analysis. In tables or charts with number, those who did not self-identify gender are counted as Not Specified (N/S).

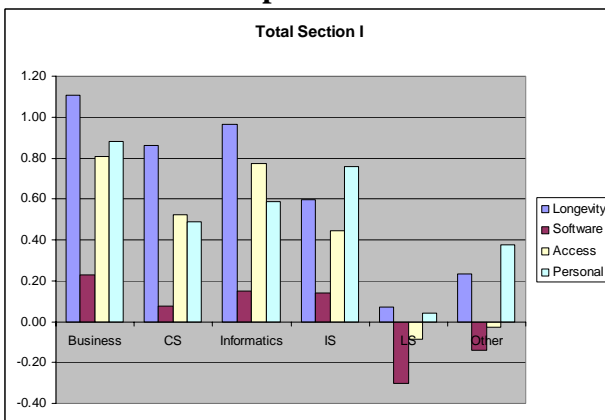
In general, the students are optimistic about the longevity of digital information and the availability of access to digital information. Being nearly neutral, the students show little

confidence in the ability of software to deal with older digital information or in business' ability to migrate their data.

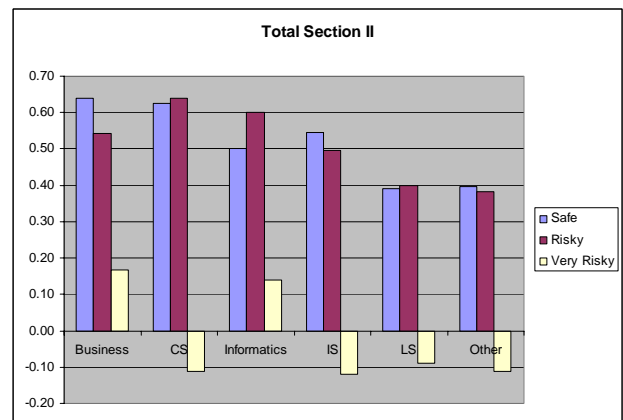
In general, the students do not have a good concept of risk in data formats. There is very little difference in the scores between the safe and risky formats. They are virtually neutral on the very risky formats.

Results by Area of Study

Overall School Optimism Factor Section I



Overall School Risk Factor Section II



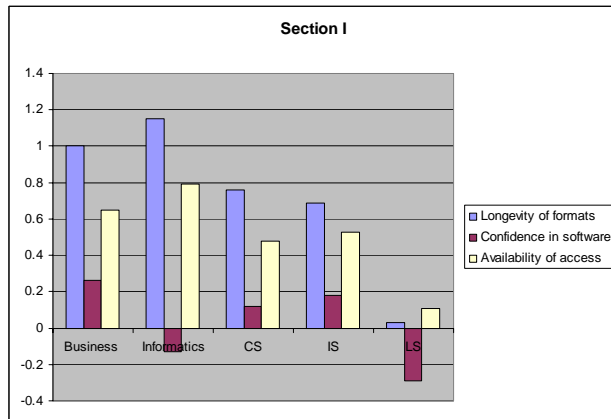
As is obvious by the chart above, attitudes towards digital preservation varies widely by area of study. Business is the most optimistic with a .76 optimism factor, followed by Informatics. Computer Science and Information Science are tied. Library Science students are the least optimistic with a negative .07 score.

Graduate Student Results

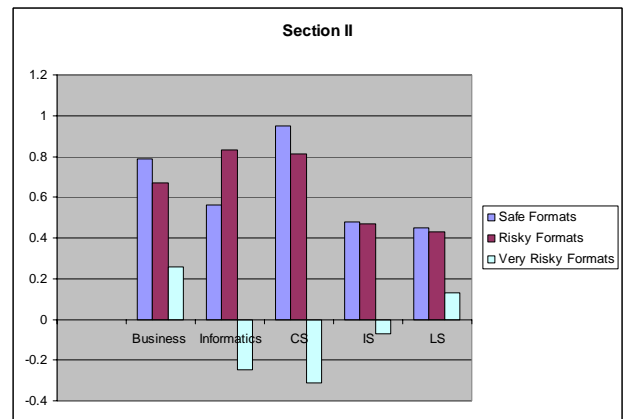
All of the areas of study have graduate programs. The general pattern of optimism is followed in the more detailed analysis. While the Business graduate students are more optimistic overall, the Informatics graduate students have the single highest optimism score; at 1.15, they are very optimistic about the longevity of digital information. For all of the graduate students, the lowest level of confidence is in the ability of software to deal with older digital information. Even

the optimistic Business students gave businesses a no-vote of confidence with their lowest score of 0.26. The LS students had the lowest score of -0.29. The Personal Practice scores indicate that the Business students have a very high level of responsibility for their digital information. LS students either do not have a personal computer and/or digital camera or do not follow good practice.

Graduate Students I



Graduate Students II



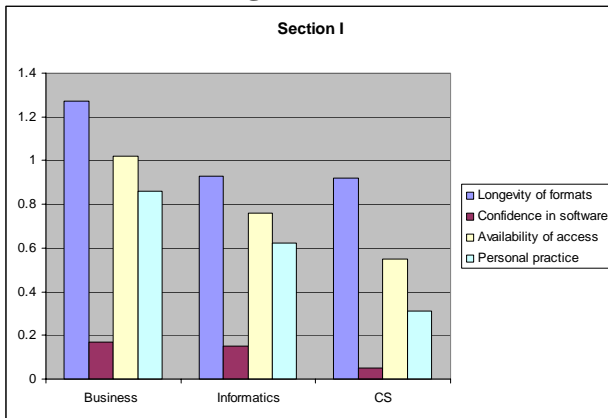
The scores for assessing the risks of digital formats follow a similar pattern with several exceptions. While business had the overall most positive scores, CS and Informatics had individually higher scores. Judging against the Library of Congress, NISO, and NARA criteria, CS was the most accurate of all of the areas of studies. They correctly assessed the highest and lowest risk formats. However, they were too optimistic on the mid-risk formats. Informatics, CS, and IS all accurately assessed the highest risk formats. LS students broke their string of lowest scores by being uncharacteristically too optimistic about the highest risk formats.

Undergraduate Student Results by School

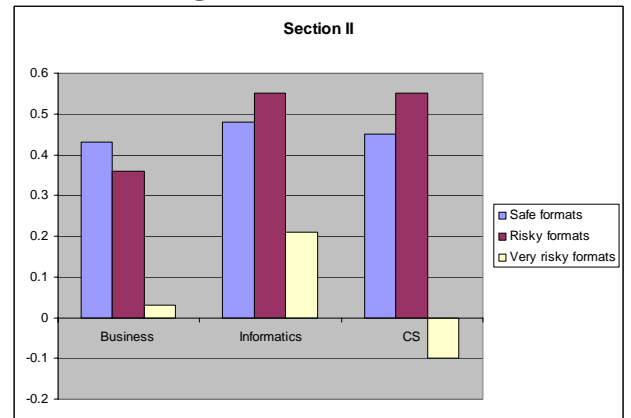
Only three of the areas of study covered by this survey have undergraduate programs – CS, Informatics, and Business. The general pattern holds. Business undergrads are the most optimistic; Informatics undergrads are in second place; and CS undergrads are the least optimistic. The

personal practice numbers are very interesting. Just over 71% of Business undergrads claim to back up their computers regularly while only 52% of CS undergrads do. The undergrads were less optimistic about formats than their graduate counterparts.

Undergraduate Students I



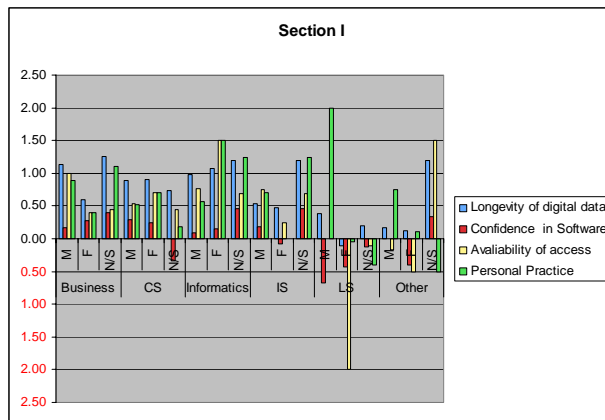
Undergraduate Students II



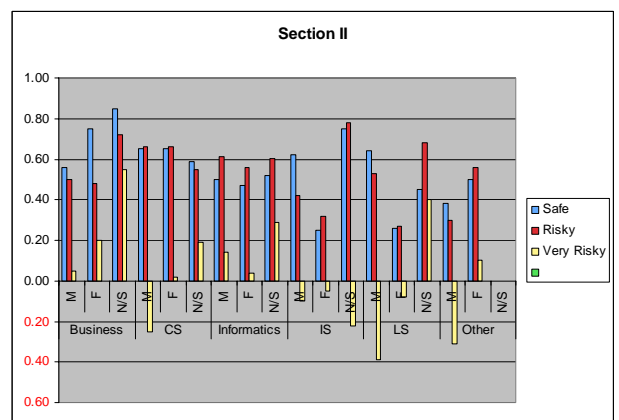
Results by Gender

As stated previously, the results by gender are incomplete, yet tantalizingly interesting. In the six areas of study that were coded – Business, CS, Informatics, IS, LS, and Other, females were either significantly more optimistic or significantly less optimistic than either males or the non-specified people in the attitudes and awareness section of the survey. In Business, LS and Other, females were less optimistic. For Other, the spread is only .02 points; but for Business, the difference is .45 points; and for LS, the point spread is a whopping .78. For CS, Informatics, and IS the females were more optimistic.

Optimism Factor by Gender by Area Section I



Optimism Factor by Gender by Area Section II



Factor Analysis

With a survey of 26 answers and 220 or more participants, the resulting dataset was large, unwieldy, and hard to decipher. Using an ‘optimism factor’ and categorizing the questions into 7 groups was an attempt to significantly reduce the data to let the information out. Another method to reduce the data was used in this study – factor analysis.

Using the statistical package SPSS, two different factor analyses were conducted. Both of the analyses used Principal Component Analysis as the extraction method. While several rotation methods were attempted, the original component matrix seemed to be the best fit to the data and the interpretation.

The first factor analysis was of the optimism factors for both sections of the survey. The table below shows the unrotated components. From this table, one can draw a number of interesting relationships between the components.

1. *There is a high correlation between the first three categories of questions in component 1. This seems reasonable. (orange)*
2. *There is a high negative correlation between the same first three questions in component 2. Again this seems reasonable. (green)*

3. *There seems to be a direct negative correction between these two components. People tended to vote in blocks – either optimistically or pessimistically. The negative correlation seems reasonable.*
4. *The personal practice question and the format risk analysis have a strong correlation in component 1. One could certainly expect that people who back up their computer regularly would know more about formats and vice versa. (pink)*
5. *In the second component, the safe and risky formats are highly correlated. Many of the study participants did not differentiate well between these two categories of formats. (yellow)*
6. *In the third component, there is a very strong negative correlation between the personal practice question and the very risky formats. One could expect that people who had good personal computing practice would know which formats were very risky. (red)*

Component Matrix

	Component		
	1	2	3
Long(1,2,3,4,10	.800	-.380	.031
Software (6,7,8)	.678	-.168	-.030
Access (3,9,10)	.814	-.397	-.050
Personal (5,11)	.414	.016	.756
Safe Formats1,4,11	.433	.769	.095
Risky(3,5,12,13,14)	.420	.799	.006
Very risky (2,6,7,8,9)	.455	.146	-.705

Extraction Method: Principal Component Analysis

a.3 components extracted.

The second factor analysis was of the entire survey results data. The principal component analysis extracted 8 components. I have used this data to test my own data reduction categories.

Category		Survey Questions
Longevity of digital information	Component 1 shows a strong correction between these questions.	<ol style="list-style-type: none"> 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	In Components 1 and 4, these questions show a strong positive correlation. In Component 3, these questions have a weak negative correlation.	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	There is a strong positive correlation in Component 1 and a negative correlation in component 3. There are positive correlations in components 5 and 8.	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	In component 6, there is a relatively strong positive correlation between these questions and negative correlations in components 3 and 8.	5. I know the technical format that my digital camera uses. 11. I back up my computer files regularly.

Category		Format
Safe formats <i>Loss-less and transparent</i>	For the safe formats, there is a weak negative correlation in component 6 and weak positive correlations in components 1, 2, and 3.	TIFF, JPG2000, ASCII, XML
Risky formats <i>Lossy and semi-transparent</i>	The risky formats have a weak positive correlation in component 1 and strong correlations in components 2 and 3.	JPG, PNG, RTF, HTML, SGML
Very risky formats <i>Lossy and opaque</i>	Very risky formats have strong correlations is components 1 and 3 and a strong negative correlation in component 2.	GIF, PhotoCD, MSWord, WordPerfect, ClarisWorks, MS Excel

Since all of the sets of questions have at least one strong correlation and many have multiple corrections, it would seem that the categories do contain related questions. See Appendix E for the complete Component table. With more work, other relationships could be garnered from this factor analysis.

Discussion

This survey was designed to discover the attitudes and knowledge that students have about digital preservation. It fulfilled its mission. A number of useful conclusions can be drawn.

Undergraduates have less knowledge and are more optimistic than graduate students. It is clear that education and experience help students know more about digital formats and the risks inherent in digital information.

Both graduate and undergraduate Business students are the most optimistic of all students. But Informatics, Computer Science and Information Science are not far behind. Library Science graduate students are only pessimistic group of students. Is this the beginning of the preservation divergence?

Library Science students learn of the history of library automation. Not only do they hear horror stories of library systems implementations and data conversions run amok, they hear of the long-term consequences to the library staff and the patrons. They learn of ancient CD jukeboxes sitting in backrooms. They learn that just as the library has paid for the newest technology, it is already obsolete. Librarians learn to distrust technology.

All of the computing science fields are optimistic. This could be because they design technology and feel some level of control over the process. They know their own motivations – to do good and to do well – to build systems that help people and to make money. When they hear horror stories about system implementations run amok, they undoubtedly assume that they could do a much better job than the incompetents running the project.

This survey can tell us much about how people perceive digital formats. With formats, familiarity equates to safety. New technologies are not trusted – even by the computing sciences students. Obviously this is a corollary with the familiarity/safety factor. But there is a significant anomaly: even though all students expressed almost no confidence in software and business in general, all of the proprietary file formats created by popular software rated high in safety. MS Word and Excel have higher safety scores than XML.

Conclusions

Virtually nothing was known about peoples' attitudes and knowledge of digital preservation. This survey shows that experience and education can raise awareness of digital preservation issues. This study also shows that people, in general, have confidence in digital data and expect that it will be available for them in the future. But they have a very no confidence that software and business will help. In order to keep the trust of their “customers”, the organizations that create, maintain and rely on this digital information have a responsibility to preserve it.

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). Retrieved on March 2, 2005 from <http://bert.lib.indiana.edu:2085/pgdweb?index=1&did=734716951&SrchMode=1&sid=1&Fmt=4&VInst=PROD&VType=PQD&ROT=309&VName=PQD&TS=1109879480&clientId=12010#fulltext>
- 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>.

Data

Research Data can be found at:

http://ella.slis.indiana.edu/~skowalczyk/L709/Survey_Data_Final.xls

http://ella.slis.indiana.edu/~skowalczyk/L709/Survey_Summary_Gender.xls

http://ella.slis.indiana.edu/~skowalczyk/L709/Total_Summary_Chart.xls

The original survey instrument can be found at

http://ella.slis.indiana.edu/~skowalczyk/Digital_Survey.doc

Appendix A

Digital Information Survey Summary Results

Section I

Questions	Total Score	Students' Optimism Score
1. Information in digital format will last longer than information on paper.	162	0.74
2. Digital photographs will last longer than film.	205	0.93
3. Academic journals that are published electronically today will still be available online in 10 years.	192	0.87
4. Digital image formats are stable and will last a long time.	157	0.72
5. I know the technical format that my digital camera uses.	160	0.73
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.	164	0.74
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.	-10	-0.05
8. Businesses migrate data from one system to another and from one record format to another with no adverse affect on their customers.	-110	-0.50
9. Information published on Websites will be accessible to researchers in 10 years.	41	0.19
10. Library and Museum digital materials (images, text, etc) are safer in electronic format.	103	0.47
11. I back up my computer files regularly.	89	0.41

Appendix A

Digital Information Survey Summary Results

Section II

The question was: If you had to store your work in a digital format and could not touch it for 10 years, which formats would you choose? The students then indicated how safe they thought that the format is for long term storage – very safe, safe, risky, very risky or no opinion

	Total Score	Students' Safety Score	Best Practice Score
1. TIFF	65	0.29	+2
2. GIF	115	0.52	-2
3. JPG	177	0.80	-1
4. JPG2000	24	0.11	+2
5. PNG	69	0.32	-1 for insufficient adoption over 6 years
6. PhotoCD	-28	-0.13	-2
7. MS Word	58	0.26	-2
8. WordPerfect	-86	-0.39	-2
9. ClarisWorks	-118	-0.54	-2
10. MS Excel	75	0.34	-2
11. ASCII	203	0.92	+2
12. RTF	92	0.40	+1
13. HTML	184	0.83	Opinions range from +1 (for content of the page) to -1 (for the context with links)
14. SGML	68	0.30	+1 (applications will deprecate over time and data should migrate to XML)
15. XML	182	0.82	+2

Appendix B

Classes Surveyed for Study Participants

Class Date	School	Course No	Level	Course Name
4/13/05	Bus	S 307	undergraduate	Data Management
4/6/05	Bus	S 518	graduate	Managing Information Security & Policy
4/14/05	Bus	S 544	graduate	Executive Leadership of IT Strategy
4/14/05	CS	A 202	undergraduate	Introduction to Programming
4/6/05	CS	B 503	graduate	Computational Complexity
4/12/05	CS	B 534	graduate	Distributed Systems
4/19/05	CS	B 434	undergraduate	Fundamentals of Computer Networks
4/19/05	CS	C 211	undergraduate	Introduction to Software Systems
4/11/05	Info	I 101	undergraduate	Introduction to Informatics
4/14/05	Info	I 300	undergraduate	Human-Computer Interaction
3/29/05	SLIS	L 401	graduate	Computer-Based Information Tools
4/19/05	SLIS	L 520	graduate	Bibliographic Access and Control
4/11/05	SLIS	L 643	graduate	Evaluation of Information Systems
4/19/05	SLIS	L509	graduate	Bibliographic Access and Control
4/14/05	SLIS	L709	graduate	Research Methods
4/22/05	CS and Info	n/a	graduate	WIC-ALL email list asking for Informatics and CS graduate students

Appendix C

Graduate Student Results by School

Section I Question Categories	Business Grads [N=29]	Informatics Grads [N=9]	CS [N=14]	IS [N=30]	LS [N=37]
Longevity of formats	1.0	1.15	0.59	0.48	0.05
Confidence in software	0.26	-0.13	0.12	0.14	-0.30
Availability of access	0.67	0.75	0.50	0.44	0.07
Personal practice	0.90	0.38	0.82	0.78	0.04

Section II Format Category	Business Grads [N=29]	Informatics [N=9]	CS [N=14]	IS [N=30]	LS [N=37]
Safe formats	0.79	0.56	0.95	0.55	0.39
Risky formats	0.68	0.83	0.81	0.49	0.40
Very risky formats	0.26	-0.25	-0.31	-0.15	-0.09

Appendix D

Undergraduate Summaries

Section I Question Categories	Business Grads [N=21]	Informatics [N=43]	CS [N=26]
Longevity of formats	1.25	0.88	0.96
Confidence in software	0.17	0.16	0.05
Availability of access	1.05	0.84	0.64
Personal practice	0.86	0.63	0.28

Section II Format Category	Business Grads [N=21]	Informatics [N=43]	CS [N=26]
Safe formats	0.43	0.49	0.47
Risky formats	0.36	0.56	0.57
Very risky formats	0.03	0.21	-0.20

Appendix E: Factor Analysis Components for all questions

	Component							
	1	2	3	4	5	6	7	8
I - 1	.606	.027	-.447	.003	-.212	-.183	.205	-.144
I - 2	.579	.007	-.454	-.120	-.327	-.138	.280	-.153
I - 3	.524	.124	-.284	-.015	.304	.081	.112	.360
I - 4	.634	.059	-.370	.130	-.215	-.009	.010	.078
I - 5	.227	.493	-.238	.027	.293	.521	-.046	-.057
I - 6	.392	.418	-.286	-.039	-.104	.146	-.081	.066
I - 7	.367	-.078	-.055	.101	.375	-.322	-.389	-.108
I - 8	.446	-.177	-.069	.210	.162	-.158	-.517	.200
I - 9	.585	-.176	-.131	-.247	.220	-.090	-.060	.379
I - 10	.545	.142	-.424	.069	-.045	-.201	.060	-.168
I - 11	.167	.244	-.242	.230	.498	.349	.200	-.197
II - 3	.386	.248	.152	.009	-.266	.349	-.407	-.218
II - 4	.386	.041	.343	.443	-.179	-.005	-.177	-.416
II - 5	.128	.359	.212	.437	-.349	.190	-.123	.371
II - 6	.389	-.277	.338	.406	-.116	-.094	.178	.336
II - 7	.534	-.297	.431	-.486	-.011	.205	.020	-.061
II - 8	.567	-.382	.417	.041	.072	.169	.144	-.078
II - 9	.441	-.454	.314	.336	.036	.085	.223	-.042
II - 10	.569	-.309	.376	-.503	-.022	.201	-.015	-.023
II - 11	.012	.642	.181	-.297	-.183	-.099	.035	.139
II - 12	.050	.591	.369	.048	-.041	.102	.221	.171
II - 13	.303	.508	.430	-.278	-.028	-.251	-.154	.008
II - 14	.155	.369	.557	.233	.268	-.252	.263	-.068
II - 15	.282	.520	.273	-.086	.233	-.342	.062	-.176