You Can Kiss Your Assets Goodbye

The state of news archives at the dawn of the digital era

Quote

"In pursuit of the bottom line, management seems to feel that it is more important to spend money on getting the paper out today than it is to archive it for the future."

Tales of woe

- **Q:** In any of your previous preservation activities (including upgrading software, moving to a new storage medium or software product) did you experience any loss of data or metadata or otherwise compromise the archives?

Kuwait Braces for Iraq's Wrath

Some citizens not as concerned

By Matthew McAllister

An Uneasy Quiet on Lebanese Border

HEZBOLLAH from A1

???

Tales of Woe

- "About 20% of stories and photos were lost when changing systems. Text is still available through the old archive. The photos are believed to be there but no one has had the time to try to get them back."

Tales of Woe

- "We had a disaster in which a vendor accidentally deleted our entire electronic photo archive. Then we discovered that the IT staffer who set up our tape backups had made an error in programming, rendering the tapes almost useless."

Tales of woe

- "We lost all the story indexing for 1987 and we don't know how."

Tales of woe

- "We are not sure why these glitches occurred. We think a previous upgrade in a different archive system might have hidden formatting, and caused the problem."

Tales of woe

- "A software upgrade caused us to lose July of 1997 for text, and a similar upgrade lost us JPEGs in our photo archive. Fortunately, these are on CDs …"

Tales of woe

"The native software that the graphics were created in has been upgraded so many times that most of these graphics are no longer accessible."

49ers pass triangle
Download FreeHand (print) | **Download Flash (Web)**
3 col x 5" diagram explains three-option triangle pass tactic used by San Francisco 49ers. San Jose Mercury 1995

Document Name	49ers pass triangle
Document Date	Jan/26/1995
Illustrator	Staff
Format	Freehand
Category	SPORTS
Keywords	sports, san, jose, 49ers, offense, pass, triangle, sports, chalk, board, mercury, 1995, football, super, bowl

Tales of woe

- "When we upgraded to Vendor A a few years ago, no one pulled the data for those photos out of Vendor B. They just dumped the photos into the new system. Hundreds of photos have no IPTC data."

Tales of woe

- "At least a year or more of text files had commas and punctuation deleted. Someone tried to fix the problem, but this resulted in commas and punctuation being put in the wrong places."

Tales of woe

- "Much of the early text files are pretty messy. It looks like some fields were either lost, or concatenated with the text."

Tales of woe

- "Very insignificant number of corrupt files ... very little -- but honestly, yes."
- "Yes, in moving to new archiving systems a small percentage of files was lost or unreadable."

Tales of woe

- "Migrating to the Web interface corrupted some files."
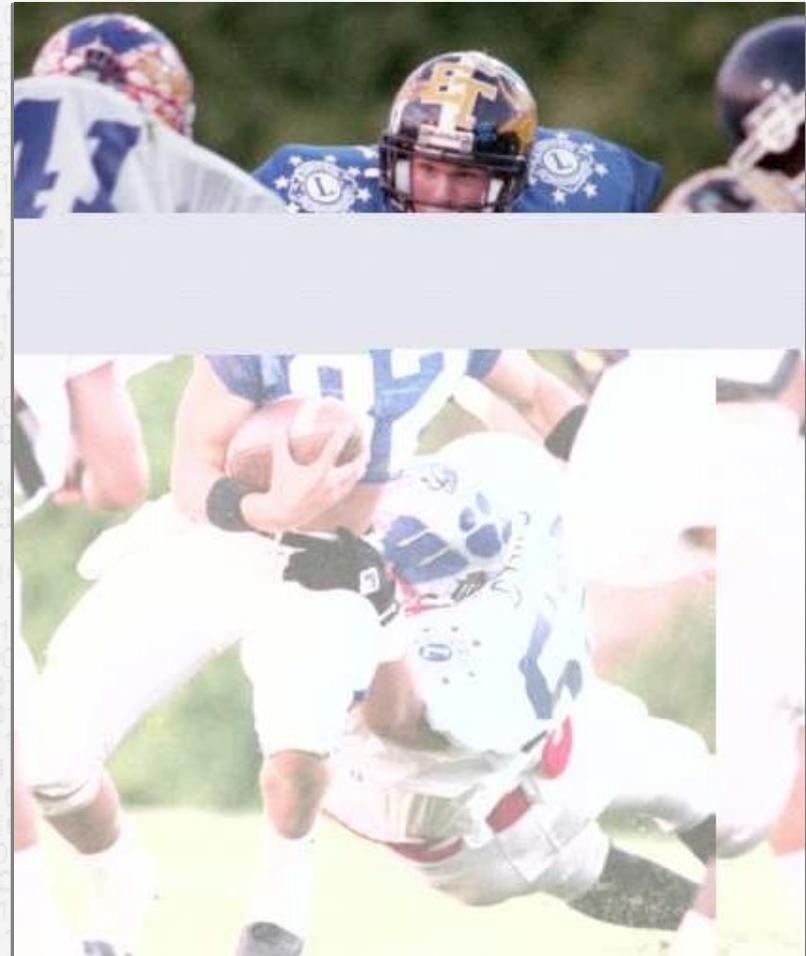
Tales of woe

- "I don't know."

Who responded to survey?

Other
11%

Newspapers
89%

How are archives used?

Chart showing (horizontal bars):
- Outside researchers: ~44%
- Corporate research: ~54%
- Sales to vendors: ~63%
- Sales to the public: ~75%
- News research: ~100%

X-axis: 0%, 20%, 40%, 60%, 80%, 100%, 120%

Size of editorial staff

>500
8%

<20
7%

20-100
21%

251-500
36%

101-250
28%
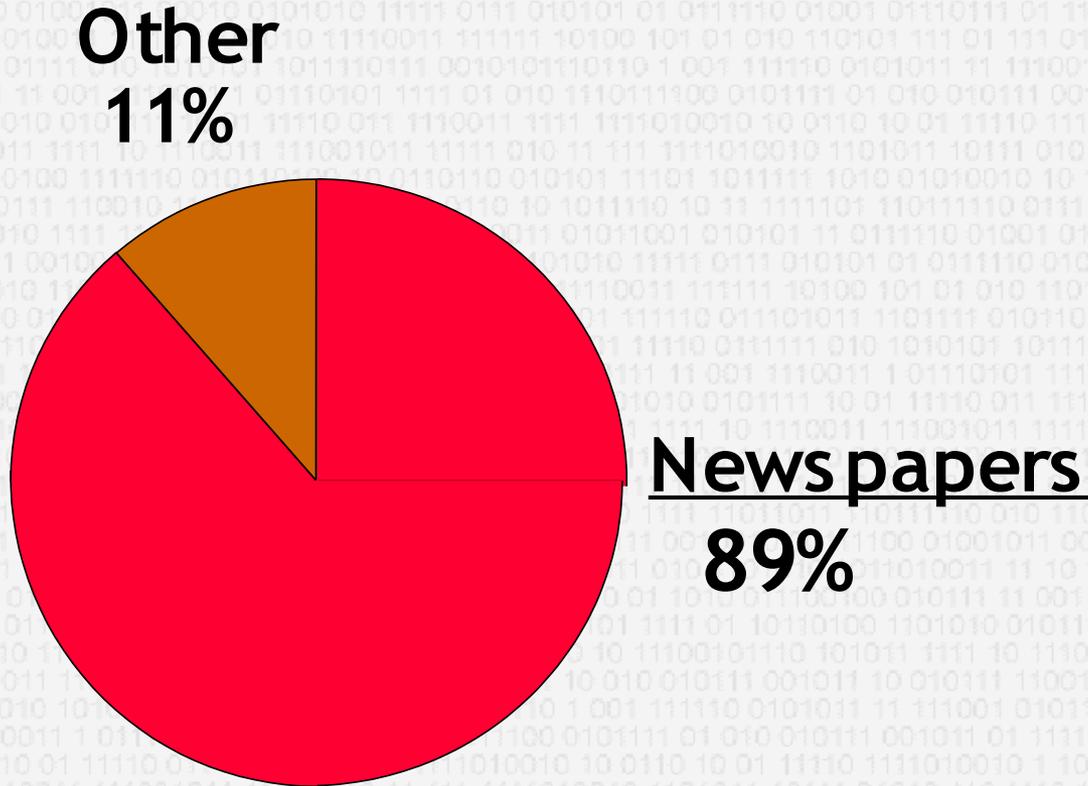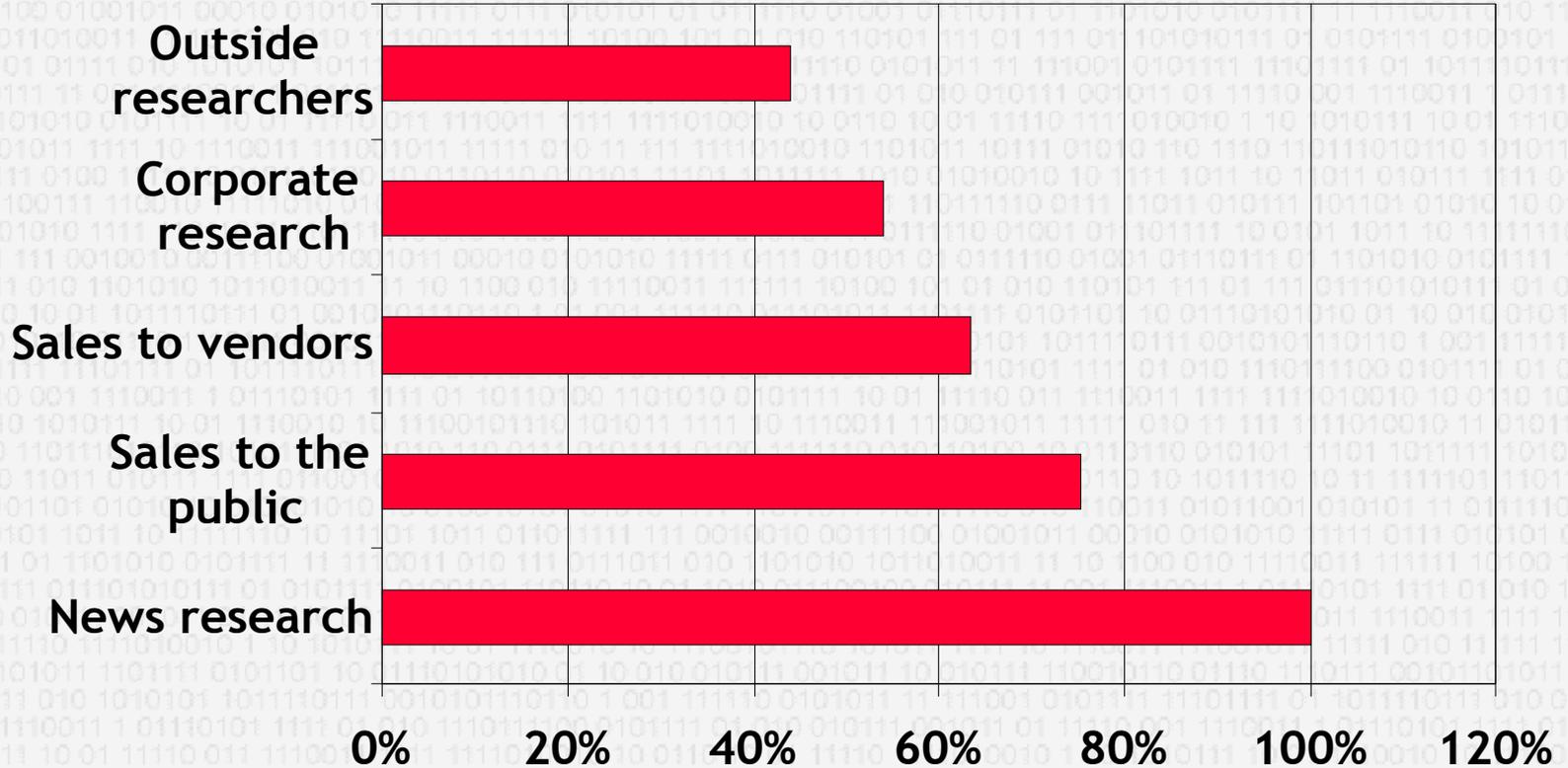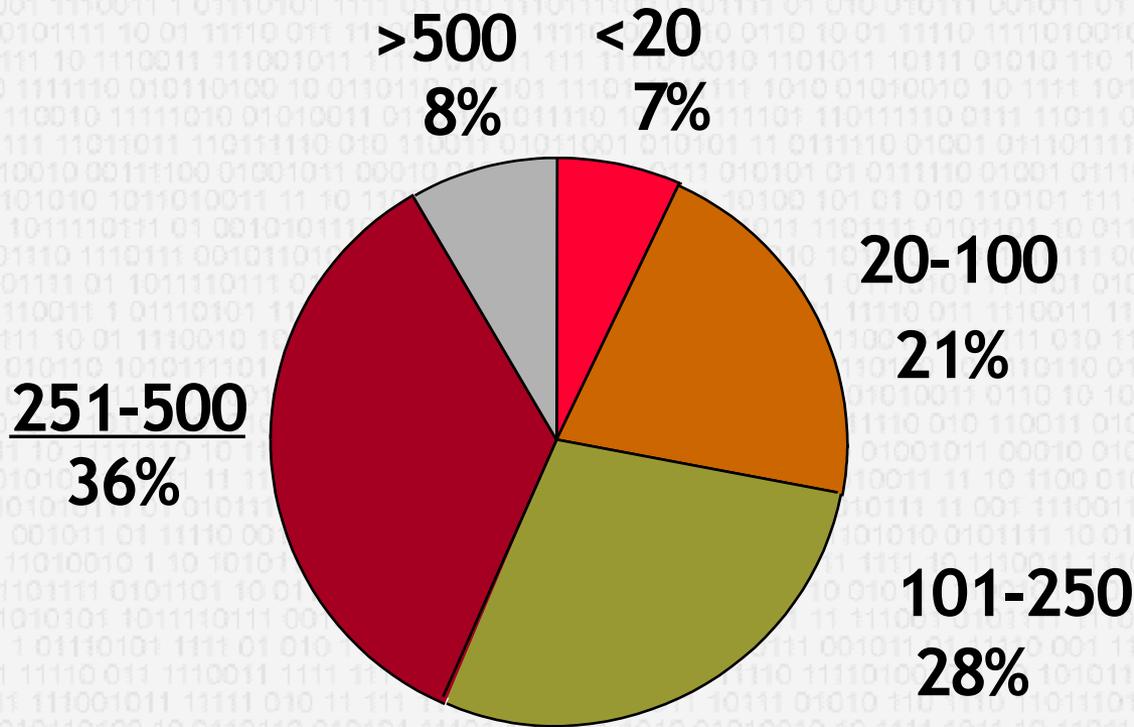
Management buy-in

Q. To what extent is your senior newsroom management committed to archival policy?

Not at all
4%

Not particularly
17%

Very
33%

Somewhat
46%
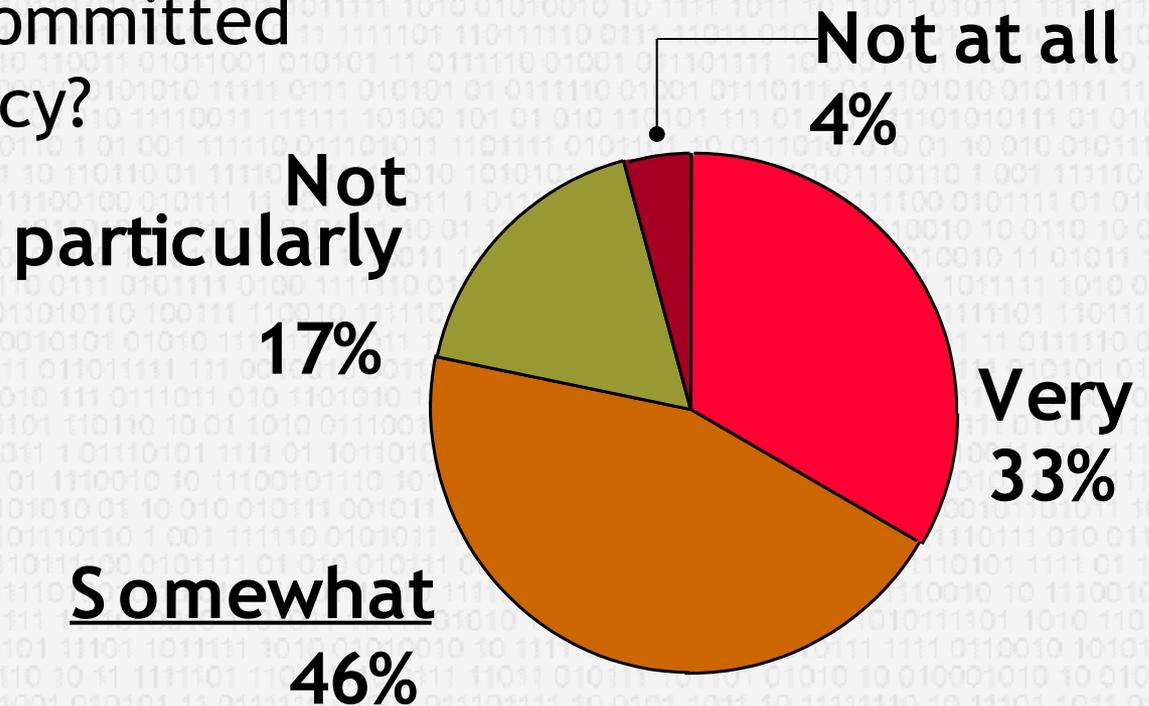
Job competencies

Q. Are there specific job competencies required for the archivist, such as professional education levels for archives management (e.g., MLS/MLIS or archives certification)?

Degree required

39%
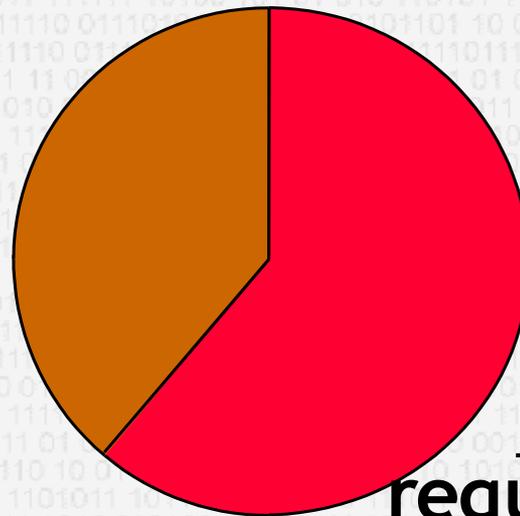
61%

No degree requirements

Job competencies

Q. Are opportunities for continuing education available to archives staff?

Yes 51%

No 49%
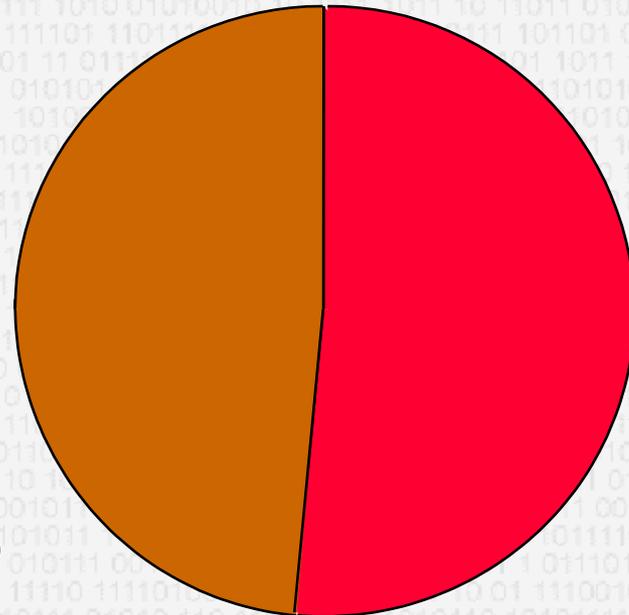
Funding

Q. Is any of the archives' budget earmarked for preservation?

No 70%

Separate budget for preservation 10%

Yes 20%
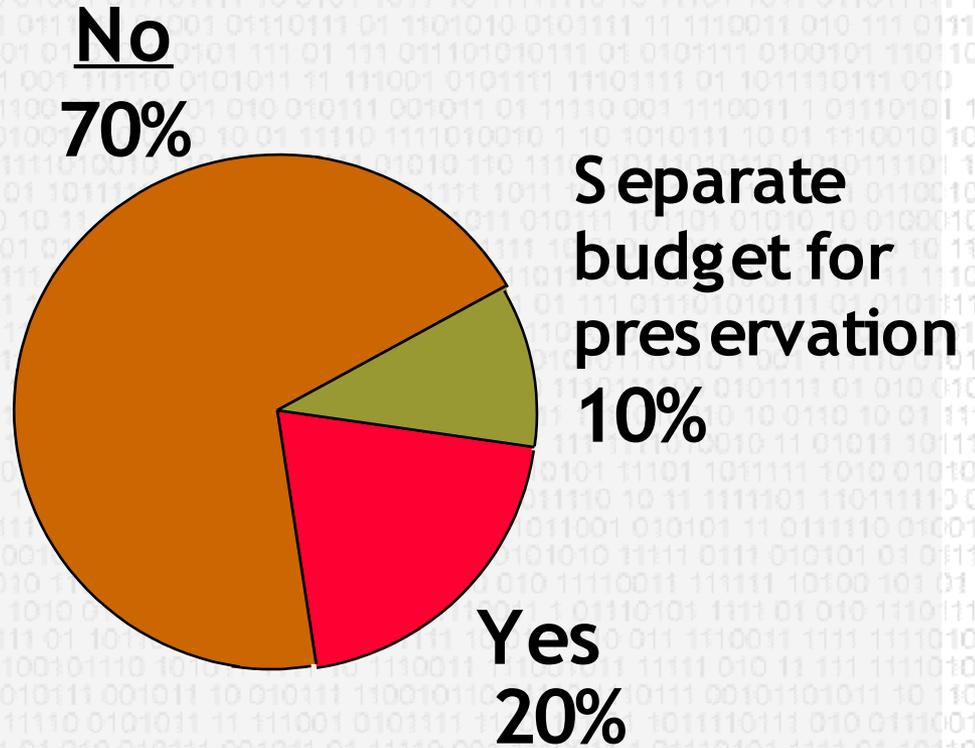
Policies for preservation

Q. Do you have a written preservation or other policy that determines what digital materials you take into your archives and for how long?

Yes 21%

<u>**No**</u>
79%
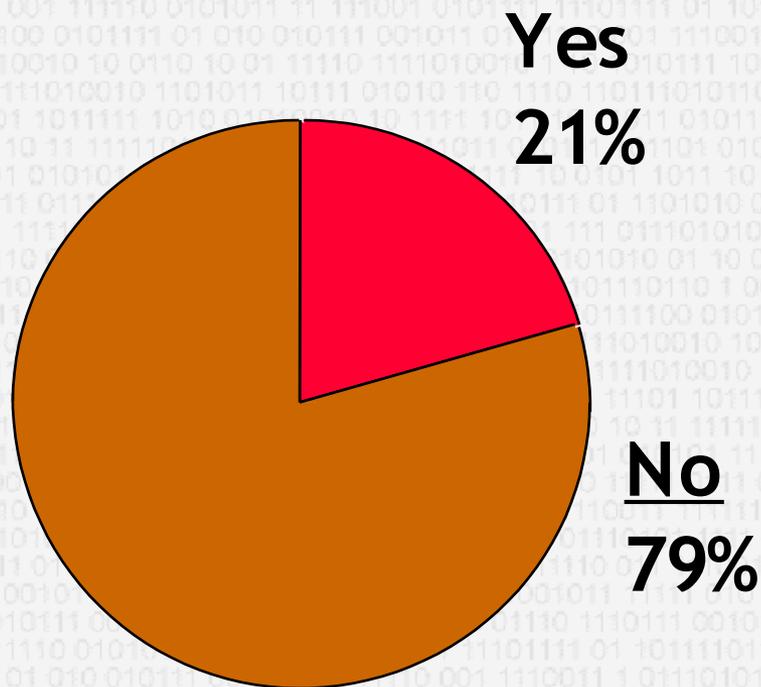
Yes, we have a policy …

- What is specified in the policy?
  - What is accessioned (87.5%)
  - Willingness to keep problematic formats (12.5%)
    - Length of time access to problem formats is guaranteed (0%)
  - Review for updates to technology (0%)
  - Digital records (83.3%)

No, we have no policy…

- Are procedures and practices familiar to staff?
  - Yes 51.6%
  - No 48.4%
- Are planning to introduce a policy?
  - Yes 40.6%
  - No 59.4%

The right to preserve

- Can you take preservation action on a file for which you do not have copyright?
  - Yes 30%
  - No 10%
  - Don't know 60%

Beyond text and photos

Q. Are you keeping web video? Info graphics? Web pages?

Video - No **85%**

Video - Yes **15%**

Graphics - No **10%**

Graphics - Yes **90%**

Web pages - No **80%**

Web pages - Yes **20%**
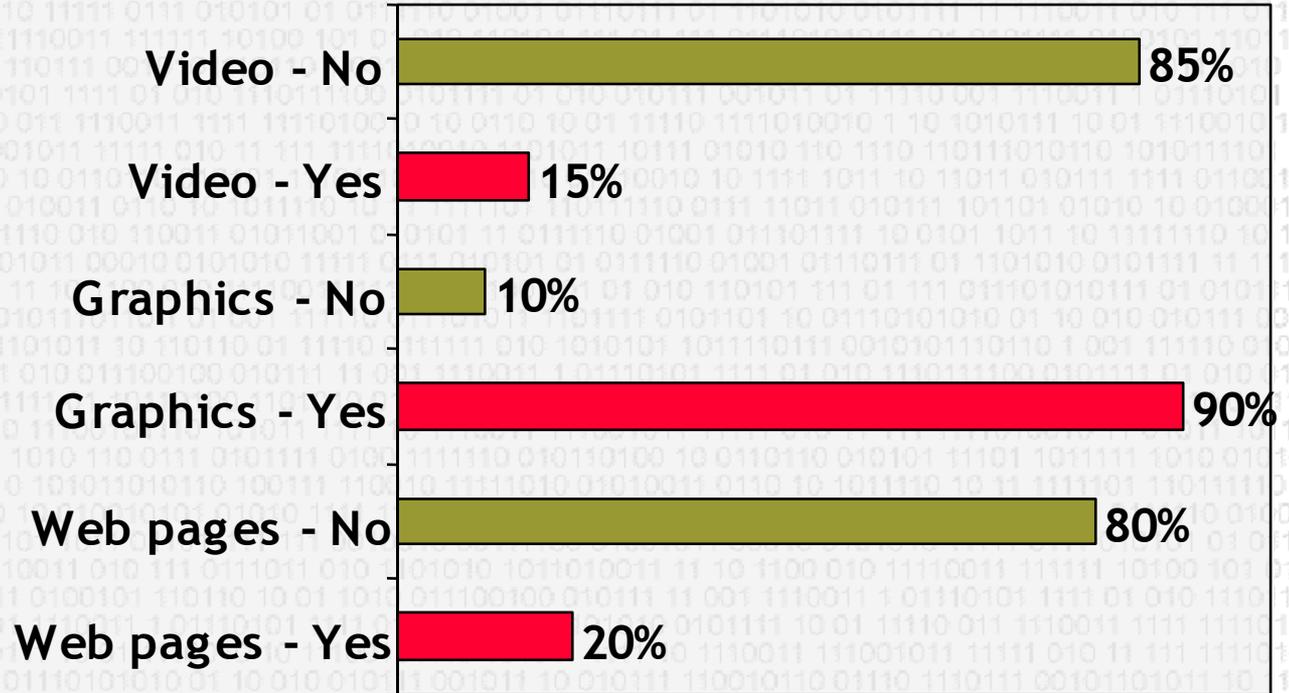
We're also starting to keep…

- Text in proprietary format **51.4%**

- Images other than JPEGs and TIFFs **48.6%**

- Spreadsheets **22.9%**

- GIS databases **5.7%**

- Specialized databases **20%**
  - FileMaker Pro, Access, Cold Fusion

Those pesky standards

- Q: Are you using metadata standards?
- A: Yes.
- Q. Which ones?
- A: I don't know

Those pesky standards

Q. Are you using a metadata standard?

No
41.7%
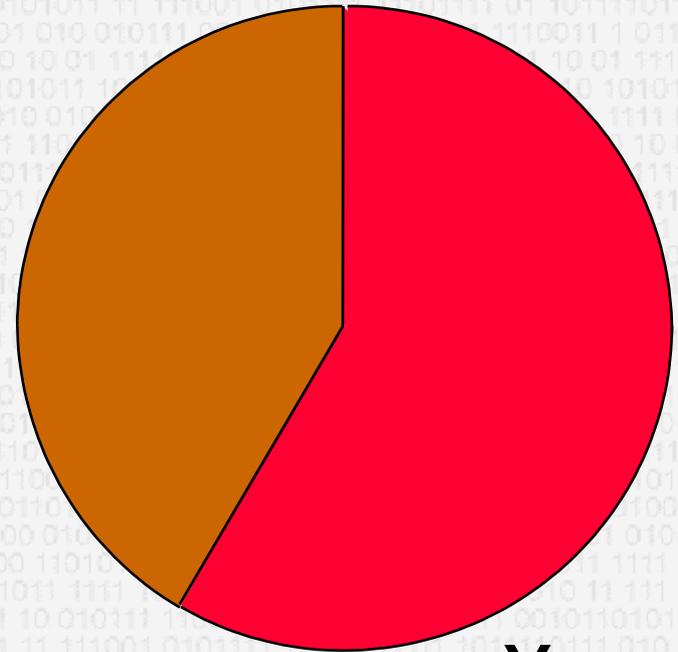
Yes
58.3%

Those pesky standards

Q. Which ones? Are you customizing them?

- IPTC
- Somewhat-to-highly customized 66%
- Vendor adaptations
- Smart answer: "We add or subtract as needed but we try to keep the fields similar so they they can flow into the next archive we get."

  ... Not a very robust standard

Those pesky standards

Q. Are you using a corporate standard for metadata?

Yes
12%

No
88%
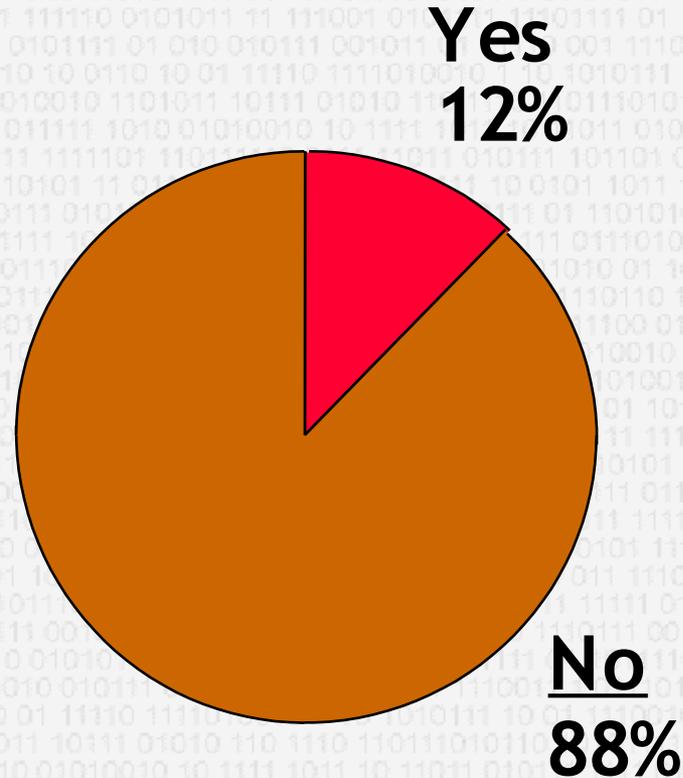
What's in your metadata?

- Rights information 69.2%
- Unique identifier 57.7%
- Software name 38.5%
- Software version 38.5%
- Hardware requirements 15.4%
- Operating system requirements 15.4%
- Peripherals for rendering 7.7%

Who's taking responsibility?

Q. Which department is responsible for hardware support?

Archives 5%

Other 13%

IT 82%
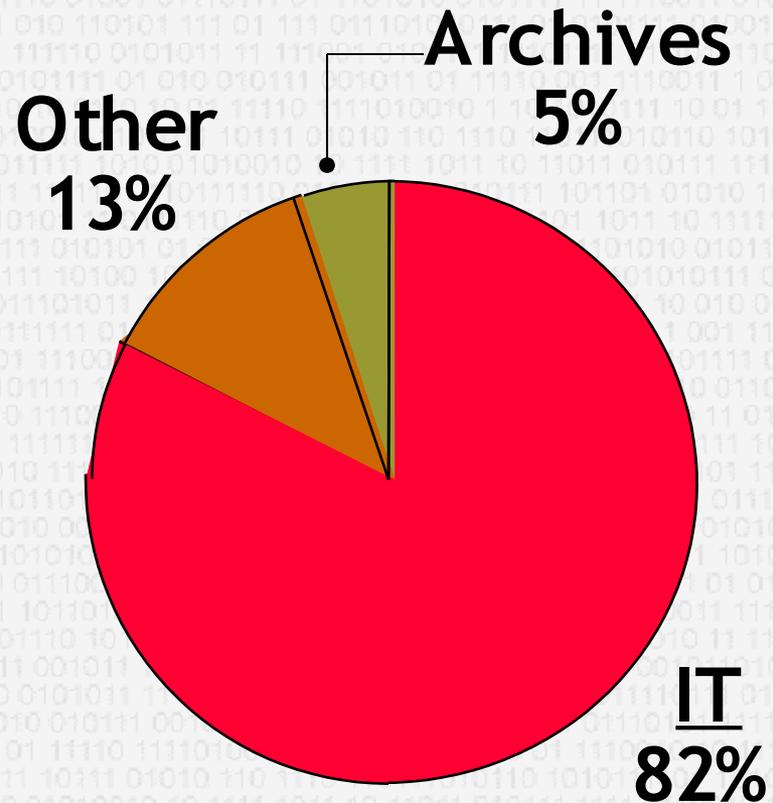
Who's taking responsibility?

Q. Which department is responsible for software support?

Archives 13%

28%

Other
-Vendors
-Photo
-Newsroom

IT 59%
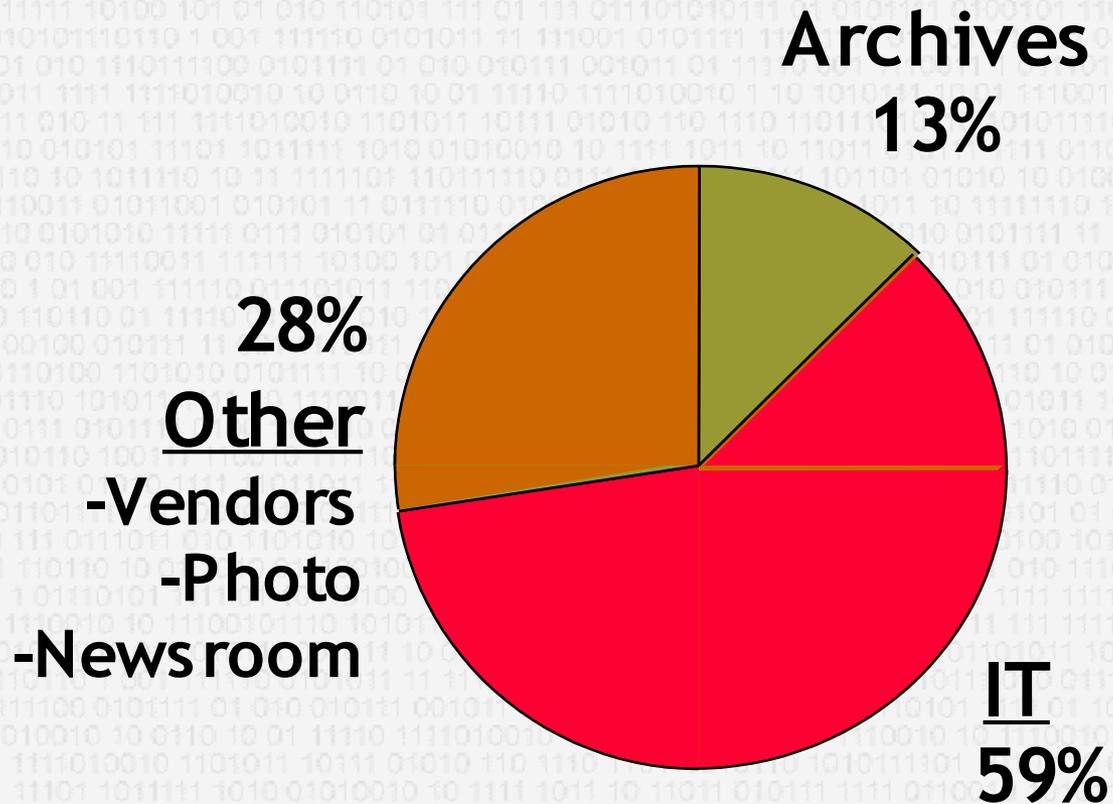
Speaking of vendors

Q. Are you currently working with a vendor on preservation?

Yes
30.3%

No
69.7%
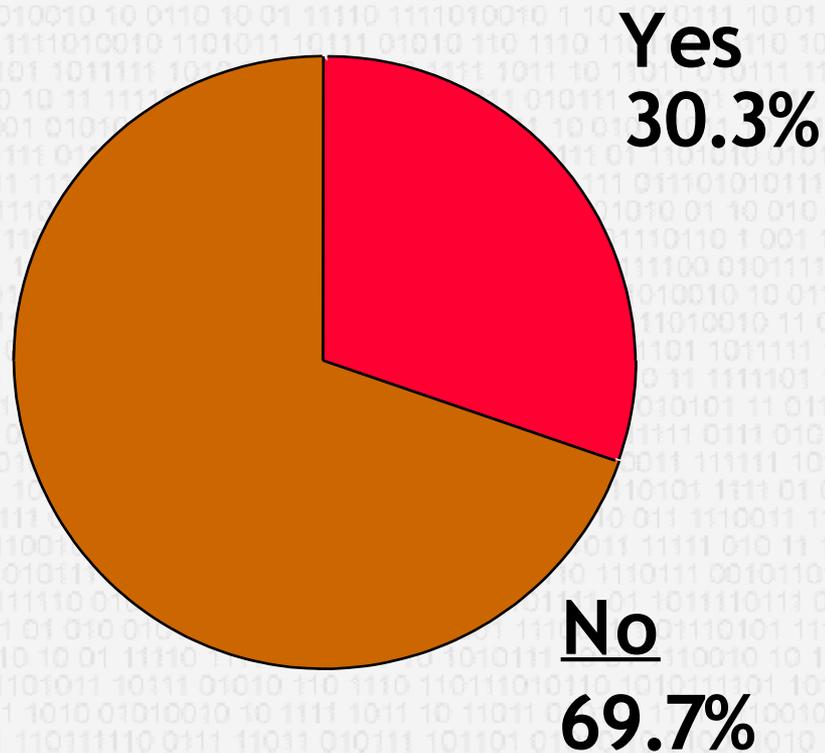
Preservation awareness

Q. How knowledgeable is your staff about preservation issues?

High 3%

Moderate 15%

Don't know 24%

Low 55%

Zero 3%
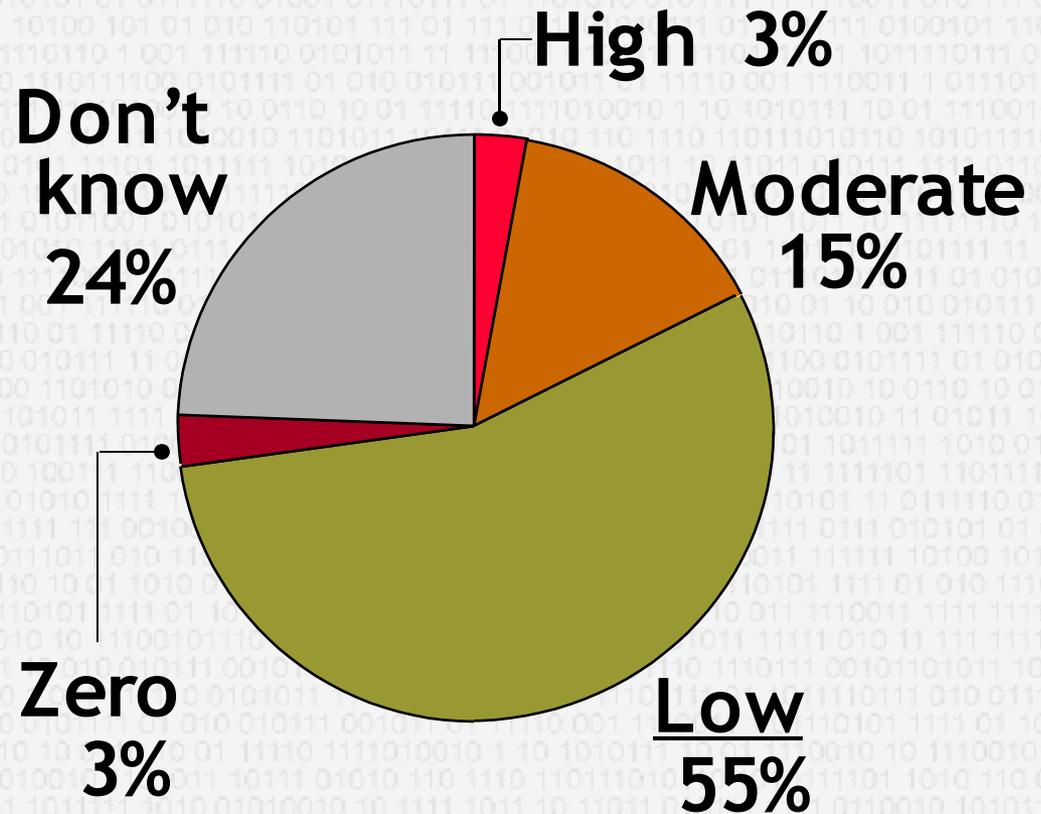
Most pressing issue?

- "Finding good preservation solutions to commit to and funding for new initiatives."

- "Arriving at standards and keeping up with the technology so that records do not become obsolete and unusable."

- "Accurate technology that makes the content completely searchable, readable and printable in the future."

Most pressing issue?

- "CD rot."
- "Changing file formats and obsolescence."
- "Getting the resources needed to do it right."
- "Need for well documented industry-wide best practices and standardization of tools and data formats."

Most pressing issue?

- "Most pressing issue is *getting* material to digital formats. We are not thinking about the preservation issues *after*."

Most pressing issue?

- "I'm afraid that once I finally have all the metadata and procedures in place for the current software, all the work could be lost with a mistake when it's updated in the future…"


Pardon Our Dust

Most pressing issue?

- "I'm afraid that once I finally have all the metadata and procedures in place for the current software, all the work could be lost with a mistake when it's updated in the future…"

42

Most pressing issue?

- "I'm afraid that once I finally have all the metadata and procedures in place for the current software, all the work could be lost with a mistake when it's updated in the future…"

Pardon Our Dust
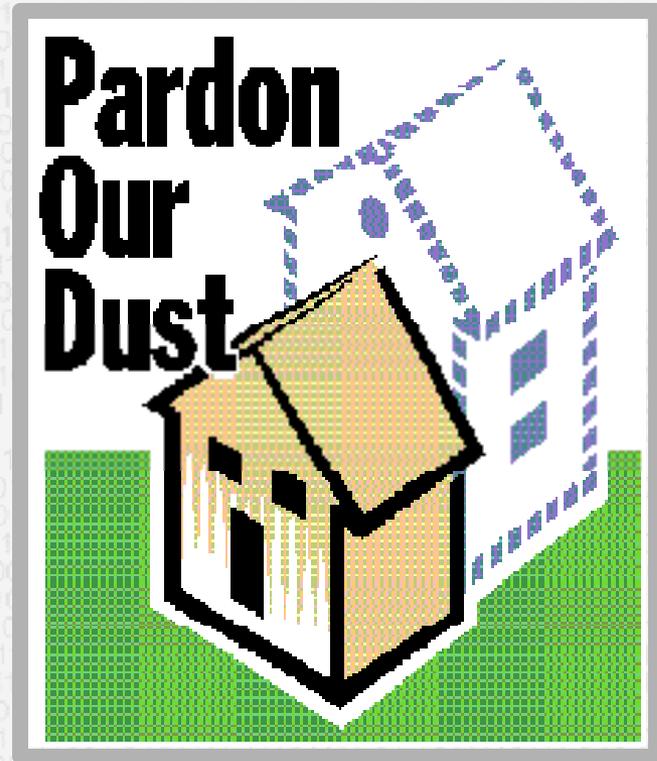
Toward best practice

- **Good news**
  - IPTC/NewsML/NITF
  - Growing awareness of the problems of digital archives
  - Continuing education available
  - Growing understanding of metadata's role
  - Rights management is robust

Toward best practice

- Needs work
  - Losses, noticed and unnoticed
  - Format control
  - Soggy standards
  - Nonexistent, weak or outdated policies
  - Resources scarce
  - Confusion over meaning of "preservation"
  - Money goes for retro digitization

Trends we're watching

- Vendor awareness

- Preservation-specific metadata

- Policy development

- Format proliferation

- More tales of woe

Next survey: 2007

No. 1 piece of advice

Keep making microfilm.

JPEG, circa 1998

Contacts

- vicky.mccargar@gmail.com
- johnsonpeterf@yahoo.com