

Archiving PDFs, Outtakes and Video

SLA Conference, June 17, 2008

Leigh Poitinger, San Jose Mercury News

lpoitinger@mercurynews.com

The Mercury News

- San Jose, California
- Daily circulation: 235,000
- Sunday circulation: 252,000
- Library staff: 3 FTEs

Our systems

- Text: SAVE, by NewsBank Media Services (www.newsbankmediaservices.com)
- Multimedia: MediaGrid, by Software Construction Company (www.swcc.com)
- Editorial system: CCI (until July 2008), but switching to Jazbox by MediaSpan (www.mediaspansoftware.com)

What's in our archive?

- Text (of course!)
- Photos (staff, wire, handouts)
- Outtakes
- Photo assignments
- Graphics
- PDF pages
- Video and photo slideshows (in progress)

A closer look at ... PDF pages

- More than 250,000 individual pages, back to February 2001.
- 40 to 200 pages per day
- At first we saved them on discs and servers, “just in case.”
- Migrated to MediaGrid archiving system in November 2005.

PDFs: The process

- Typeset page generates a PDF, which is automatically exported into MediaGrid
- Metadata included: page number, section letter, date published, day of the week.
- All of the text on the page is indexed and searchable
- Archivists add section name
- Daily time commitment: 5 minutes

PDFs: Benefits

- Fully searchable & easily accessible
- Preserves exact look & feel of page
- Low time commitment
- Stable file format, open specs, will be around for years to come
- Potential revenue benefit

PDFs: Issues

- Still need to microfilm
- Are we getting every edition & page change?
- Single pages, not linked to others
- Storage requirements
- Metadata on older PDFs incomplete

PDFs: Future plans

- Link PDF pages to text & photo archives
- Link story jumps
- Link corrections?

- **BOTTOM LINE: Just do it!**

A closer look at ... Outtakes

- We've been saving digital outtakes since January 2007
- A typical photo assignment can generate 10-200 outtakes (depending on the photographer)
- 500,000+ outtakes and counting!

Outtakes: The old process

- Photographers drag & drop all outtakes from an assignment into MediaGrid
- Linked by assignment number
- Searchable by assignment number, caption info, photographer name
- High-resolution images stored on MediaGrid
- But ... this takes up 1-5 terabytes of data per year

Outtakes: New solution

- Low-resolution file to MediaGrid for searching
- High-resolution file to a Blu-Ray drive
- High-res images are then burned to Blu-Ray discs (each disc can hold 25-50 GB)
- Still searchable by assignment number and caption info

Outtakes: Costs

- Set-up cost: \$23,000
- Included: installation, printer, Blu-Ray burner, server & database software (DAX Archiving Solutions, www.daxarchiving.com)
- Ongoing cost: discs (\$10-\$12 each)
- More on Blu-Ray: www.blu-ray.com/info/

Outtakes: Issues

- Is it worth it?
- Still have to manually retrieve the discs & store them somewhere
- **THE BOTTOM LINE:**
If you can afford it, why not?

A closer look at ... Video

- We've been doing audio/photo slideshows and videos since 2005
- We're producing lots of video ... the paper's goal is five videos per day
- Web hosting using Maven Networks, but switching to Brightcove soon

Video cont.

- Alphabet soup of files:
 .mov, .flv, .mp3., .jpg
- Photographers shoot most stuff in high-definition; reporters shoot non-HD
- Typical quick digital video project is about 1.5 GB
- De-facto archive on Maven, and scattered external FireWire hard drives

Video: Issues

- Is there a business case for archiving video? No one has monetized it yet.
- What to save: raw material, finished projects, everything? Who decides?
- Compatibility issues and standards.
- Metadata: how much, who will do it, quality control, staff time needed?

Video: Planning

- Figure out what & how much you are producing, and estimate storage needs.
- Decide where & how to store the data – does your archiving system handle multimedia?
- Create a retention policy.

Video: Planning, cont.

- Create a workflow process.
- Create a metadata scheme.
- Accept imperfection.

Video: Our process

GOAL: Move finished projects into MediaGrid for indexing and searching, and put selected raw material onto Blu-Ray

Video: Our process, cont.

Same workflow as photos & outtakes:

1. Fill out assignment form & get number
2. Upload finished project into MediaGrid
3. Raw material gets copied to Blu-Ray
4. Linked by assignment number

Video: Our progress

- Assignment form issues
- Staff time
- Metadata scheme
- No IPTC fields for video

Video: Metadata

- Slug
- Filename
- Byline
- Byline title
- Source
- Credit
- Copyright
- Summary
- Section (news, features, etc.)
- File size (ex: 1.3 GB)
- City
- State
- Country
- Date created
- Date published
- Software format (ex. Flash v.6)
- File type
- Assignment number
- Length (minutes, hours)
- Headline
- Keywords
- Status (published, archived, etc.)
- Special instructions

Video: What's next?

- Refine metadata
- Deal with workflow issues
- Tweak retention policy
- **THE BOTTOM LINE:**
We're still figuring it out. Stay tuned!