

# What's New in Archiving

"It's Not Your Father's Archive Anymore"

Generously sponsored by Preservati Consulting

```
01101010111 0011100011101
001011000110110011100110001
001100 1011110010 1110101
0100010 100101100110011011
1 110110 100 011001011010
00 110010010110010101001100
10111 001100111001111001010
```

Preservati Consulting

SLA News Division  
SLA Annual Conference –  
Seattle 2008

# What's new?

- Blogs, RSS feeds, Wikis, other unique online content – are you archiving these?
- XML components in archive systems allow us to link story texts to PDFs, photos, audio, video
- Taxonomies (not new, but more and more evident on web sites for SEO)

# What's new?

- Print organizations producing and archiving video, audio
- Web scraping
- Extracting data from PDFs
- Librarians embedded within online teams
- New ways to handle mountains of photo outtakes

# Chicago Tribune Data

- Newsroom production system: CCI v6 (looking at NewsGate)
- Text archive: SAVE system (NewsBank)
- Photo archive: Merlin

# Archive initiatives

- Addition of new taxonomy module from Nexis to be incorporated on our front end before stories are transmitted to vendors
- Spanish language archive of our Hoy publication
- Blog content, other unique online content
- Text archive system supported on SAN rather than on separate box
- Special archive collections for use on our web site

# Orange County Register

- Garbage in, garbage out.
- CCI's NewsGate product has "Long Term Storage" function not expected to be an archive solution. OCR would like to see a move to make this an archive solution.
- OCR has added news archive material from 1987-2005 to the data that their NewsGate system can search across. Search functionality needs further development work.
- NewsGate fields called Contextual Use and Category can be used to populate fields in the news archive. This relies upon reporter input of metadata at the front end and there typically is not full compliance.

# Providence Journal

- Providence Journal stories move directly from CCI to vendors without human intervention (mostly) after the feed manifest is checked for transmission. Problems: how to handle on weekend, missing headlines, duplicate captions, unlinked elements (sidebars etc).
- The feed into the internal archive (SCC) is also fully automated, although we are still touching every story to make sure that it is in the archive with the proper metadata. Staff members spend about 4 hours per day enhancing stories. Still some manual work cutting and pasting in recipes and sidebars, and quality control work.
- Process for photo and graphic material to flow into SCC with CCI publication data is not yet working so publication and keyword info is added manually.

## Session Speakers

- Debra Bade, Chicago Tribune
- Leigh Poitinger, San Jose Mercury News
- Ron Larson, Wisconsin State Journal/Capital Times
- Chuck Palsho, NewsBank
- Katie Daugert, National Public Radio
- David Breslauer, MerlinOne



# What's New in Archiving

“It's Not Your Father's Archive Anymore”

Generously sponsored by Preservati Consulting

```
01101010111 0011100011101  
001011000110110011100110001  
001100 1011110010 1110101  
0100010 100101100110011011  
1 110110 100 011001011010  
00 110010010110010101001100  
10111 001100111001111001010
```

Preservati Consulting

SLA News Division  
SLA Annual Conference –  
Seattle 2008