

“On the Record”
A Forum on
Electronic Media and the Preservation of News
October 23, 2008
New York Public Library, Celeste Bartos Forum

Towards a Digital Cooperative: How the Associated Press Distributes and Archives the News
Valerie S. Komor, Director, AP Corporate Archives

The world’s oldest and largest news organization, the Associated Press, was founded in New York City in 1846 as a cooperative, with American newspapers as its owners and members. Today, it is still a not-for-profit cooperative and, with the U.S. Post Office, the only 19th century national communications system to survive in its original form. Western Union sent its last telegram in January 2006 and is now just a financial services company. How AP has managed to retain its structure and still preside at the nexus of global journalism for 162 years is a story that remains to be told. Today, I will talk about AP’s rapid transformation from a cooperative linked by metal wires to a cooperative linked by digital networks. I’ll also review how news delivery and news preservation is affected by these changes.

Before the invention of the telegraph in 1844, news traveled as quickly as the swiftest horse, steamboat, or carrier pigeon. The “news” that traveled was foreign news, printed in foreign newspapers and collected by newspaper agents in rowboats who met incoming transatlantic ships at Halifax, Boston, and New York. Regional news might include election results or presidential messages and congressional proceedings from Washington. All other news was locally obtained.

Ingenious measures were taken to speed this news to its destination ahead of the competition: hence the preoccupation with speed that permeates journalism lore. In 1843, for example, the printers of the *New York Sun* waited aboard a steamboat docked at Piermont on the

Hudson for the text of the governor's New Year's message. When it arrived by railroad, the printers set up type on the boat as they headed down river, allowing the Sun to proclaim the next day:

“By the Sun's Exclusive Express. From Albany Through by Horse and Sleigh in 10 Hours and ½.”

During the spring of 1846, as the telegraph expanded up and down the East Coast, linking New York and Washington on June 5, the second publisher of the *Sun*, Moses Yale Beach, had the idea of combining the horse with the U.S. mail coach in order to speed Mexican War news to the nearest telegraph head near Richmond. The dispatches originated in Veracruz, crossed the Gulf of Mexico by boat, and landed at Mobile, Alabama, where they encountered delay in waiting for the Great Southern Mail. Beach employed a pony express rider to carry the dispatches ahead of the mail to Montgomery. There, they rejoined the mail for the journey to Richmond and were put on the wire to Washington and New York. Beach did not pay his riders unless they gained a 24-hour edge over the mail, which they mostly did.

Beach's innovation was to offer an equal share in the pony venture to the New York daily papers. An inventor with a list of patents to his name, Beach doubtless understood that the telegraph would soon replace the pony and railroad expresses. He must also have seen that the telegraph, the first “through” instantaneous communication with New York at its center, removed the impulse to compete in order to transmit the news. In fact, the telegraph would require cooperation to avoid duplication of effort at the telegraph office, because the offices limited each sender to 15 minutes. Why wait in line to send a common digest of European news, when the next person was going to be sending the same thing? Not to mention that what content could be sent in 15 minutes was extremely limited. An association of newspapers was formed not because the telegraph was more expensive than ponies; it wasn't. It was formed out of the self-interest of the parties, who recognized they were better off cooperating than competing. Thus was born the Associated Press of New York.

The timely combination of Morse's telegraph and Beach's news service signaled a revolution in the delivery of news that has only been equaled by the digital era. For the first time, strangers living far distant from each other could acquire the same information at the same moment. Editors could collect news as it was breaking rather than rely on previously published reports. As the telegraph spread nationwide, AP responded by establishing a system for selling telegraphic news and inviting American newspapers to share the cost of the service. The AP operated like a gristmill, processing the national and international input of news and producing a uniform output. Communications scholar Menahem Blondheim suggests that AP's peculiar strengths helped shape an American national identity in the second half of the nineteenth century.

[AP's] structure as a national institution—impersonal, non-local, unselfconscious, and hidden—gave wire service news, however partisan, the appearance of objectivity. The Associated Press helped Americans accommodate to a common information environment. By giving news that impressed the minds of Americans a national orientation, it fostered the integration of American society.

While AP's mission—"to gather with economy and efficiency an accurate and impartial report of the news"—has not changed in the last 162 years, the digital revolution is remaking the way AP delivers the news. Instead of giving the same thing to everyone at the same time, it now gives people what they want when they want it. The "common" information environment that prevailed from the last half of the nineteenth century through the late twentieth, facilitated by the ubiquity of the daily newspaper, has yielded to proliferating and personalized information worlds, some no bigger than the palm of the hand.

At AP, this transformation has been underway since 2003, when Thomas Curley arrived as President and CEO. He introduced changes in AP business systems to make AP a more agile player in a highly competitive media environment. His signal accomplishment to date has been to retool the cooperative without altering its purpose. He has moved AP from the traditional wire service, which delivered news in one great stream, to a self-service operation in which members access AP content through various web portals and services. What happens to the content within AP is also changing. Where we once took stories in, processed them, and moved them out in more or less the same format (text or photo or graphic), we are now aggregating or collecting content in multiple formats and simultaneously channeling it out through numerous pathways. The aim is to capture a new generation of consumers who are as fragmented as they are selective. AP's recent award of its 2008 Chairman's Prize to the Mobile News Network attests to its desire to reward entrepreneurship directed towards the end user, whereas AP has traditionally been a wholesaler of news. The Mobile News Network delivers news from AP and its members to smart phones and allows users to customize features for favorite sections and locations, view video and photo galleries, search a 30-day archive, and share content with friends. None of this would be possible were not 1,000 American newspapers sending AP their content for enhancement, enrichment, and normalization. This is the digital cooperative in action.

Before going into greater detail about how AP distributes and archives its news, I will survey how AP has done so in the past, when newspapers and broadcasters were its primary clientele. My focus will be on the print report. Note that much of what I say about AP's historic operations still holds true today.

Drawing from reporting across the United States and in 97 foreign countries, AP has long provided text to newspapers and broadcasters that they are free to cut, expand, or combine with information from other sources, or otherwise rewrite to highlight local angles. What AP puts on the

wires, then, is the only record of what AP has written, and much of what is published or heard is never credited to AP as the source. Therefore, AP reports are distinct from the published account and possess a corresponding intrinsic and evidential value. In paper format, these reports have included hand-written dispatches (now extremely scarce), telegraph cables (1846—ca. 1914), and wire copy off the teletype (1914-85). Even the electronic text does not duplicate what a newspaper or website publishes.

For most of its history, AP covered the news in cycles based on the existence of morning and afternoon newspapers. This distinction began to disappear in the digital age, when transmission speeds and news volume increased dramatically, causing papers to extend publication hours to a virtual 24-hour cycle. During the “AMs and PMs” era, the AP wire began moving copy at noon for morning papers whose first editions actually came off the presses that evening (with next day's date) and continued until after midnight. The PMs cycle opened at midnight or 1 a.m. for afternoon papers, whose first editions came off the presses a couple of hours before noon and continued into late afternoon. All this reflects the fact that most news originated during daylight hours and might or might not be carried over into the next day or beyond.

Stories originated with the reporter's written draft or dictation from the field, were edited for the wires and filed. At AP, there have been many wires: the A or national wire; the B, a secondary trunk wire; the E (a regional East wire); the G (a regional South wire); the F or Financial wire, the S or Sports wire, and fifty state wires. The international wires were the Latin American and Tangiers-Pacific, a world wire linking New York via Tangiers to Europe, Africa, and the Middle East, and New York via Tokyo to Asia. Some of this terminology persists, especially reference to the A, F, and S wires, even into the digital age.

Updates during a news cycle are called “leads.” The process of adding details to the story as they become known may go on for several hours or days, but every addition is filed on the wire. The story of Congress' struggle to pass a 700 billion dollar bailout package for Wall Street required 52

leads, from 3:12 a.m. September 29, 2008 to 10:03 p.m. that night, a number surpassed only on Election Night.

The continuous versioning of the story in real time is what makes AP journalism unique, and the electronic news production system, in use since 1996, preserves the versioning function. Reporters compose their stories using an editing tool known as Reporters Workbench, an interface client program that resides on the PC and accesses the production system. At the same time, reporters can access the Text Archive, a central digital repository containing national, state, and international reports dating back to 1984. When an editor sends a story to the wire, it automatically goes into the Text Archive, together with any internal messages, story lists, and advisories not intended for publication.

Even as we speak, vast changes in content management and distribution are underway, made possible by the development of “eAP,” or the “electronic AP.” Announced in 2003, eAP is a multi-faceted initiative that is making news delivery an interactive operation between AP and its members and customers. A content enrichment and distribution platform, eAP delivers news in all formats—text, photos, graphics, audio, video, and interactives—in packages members can easily use. The eAP system offers access to this content from multiple AP news and market wires, satellite channels, historical archives (especially photo, film, and video) and other sources. At the heart of eAP is a database which holds all published content. AP Exchange is the web portal that allows regulated access to this database, and all web portals and distribution systems work off of the same database. In the near future, AP intends to provide the industry with software tools that will allow broad and creative access to this important historical repository.

Underpinning eAP’s search capability is a content enrichment process, which is the primary calling card of the “digital cooperative,” the initiative charged with finding new digital outlets for the

news produced by AP members. AP is now able to tag its own and member content with a robust set of metadata that will trigger hits from search engines, customize content, and support contextual advertising. The goal is to enable AP and member content to more easily float to the top of search engine results. A team of information specialists is developing this proprietary classification service using the Associated Press Publication Language (APPL) schema, standard taxonomies, and auto-categorization rules that will let reporters and editors apply a structured set of subject terms to stories prior to filing. When a story is filed, it undergoes auto-categorization and is piped into eAP, where members can view the content and search within content “verticals” such as Sports, Entertainment, Business, National Report, Environment, and hundreds of other subjects. Eventually, eAP will replace the Text Archive as the central news repository. Staff are now completing the lengthy process of extracting the more than 65 million documents from the Archive and ingesting them into eAP.

What I'd like to do now is show you how a major financial and national story developed, and some of the channels that it passed through after it was filed. This is the September 2008 story of the government's 85 million dollar bailout of the insurance giant American International Group, or AIG. AP's “Money Team,” a group of eight reporters and one editor, worked the story, which was first reported from Charlotte, North Carolina, then from New York, and finally from Washington, D.C. AP business writer Ieva Augstums, who covers insurance from the AP's Charlotte, North Carolina bureau, was called into the office on Sunday, September 14. She had been following the restructuring of AIG and the fall in its stock price. We start with her first story:

- September 14, 7:50 p.m. Second version of the story (first was First Lead). Augstums reports from Charlotte that the *Wall St. Journal* is reporting that AIG plans to disclose restructuring plans by early Monday. The story is moving only on the F (financial) wire and APFN, the premier financial services wire.

- September 15, 12:35 p.m., the AP Business Desk in New York files the first lead of the day in the AIG story (now slugged BC-AIG), incorporating Augstums' previous reporting. The story is a breaking news story ("Breaking:True) or NewsNow, and is routed on the A or national wire, the F or financial wire, the premium APFN service, and online. News Now means it is written so that broadcasters can read it. Says New York Governor David Paterson will allow AIG to use \$20 billion in subsidiary assets to provide cash to stay in business. Shares of AIG had fallen 51% in afternoon trading.
- September 16, 6:44 p.m. 11th Lead. Business Desk in New York files a Source Report ("A person with knowledge of the situation?"). New Headline. NewsNow. Adds the required bylines for source reports (Ieva M. Augstums in Charlotte, Stephen Bernard in New York). There are two new headlines: the first under 50 characters; the second under 94. Editor notes the story "will be led" or added to as more information becomes available. You can link to photos and video. Ticker symbols are highlighted for the use of financial websites like Yahoo Finance, which will show these stocks and link to their performance, and pull stories relating to these companies.
- September 16, 6:49 p.m. 12th Lead Writethru. The print version of the previous. Same headline, but with updates and more details. The lead is reversed from previous ("The government is increasingly likely to step in to help rescue the huge insurer AIG, a person with knowledge of the situation said Tuesday.")
- September 16, 8:02 p.m. 16th Lead. New Headline. NewsNow. The dateline switches to Washington because the Federal Reserve is making an announcement that the U.S.

government has agreed to provide an 85 billion emergency loan to AIG. Broadcasters interrupted their programming to read the report.

- September 16, 8:24 p.m. 18th Lead Writethru. First bylined story from Washington. Adds Jeannine Aversa (Washington) to Augstums (Charlotte) and Bernard (New York). Editor notes: ADDS more details, background, restores graphics, bylines. Aversa takes over the story from Washington.
- September 17, 1:51 a.m. This is the overnight story, recast with a softer, more analytical approach. Note the lead: “Another day, another bailout.” There was no more hard news to require a new headline. Morning papers would have taken this story.
- September 17, 7:20 p.m. This is the final story of the day, by Jeanne Aversa, with a lead that sets the story on Wednesday and recasts it: “American taxpayers awoke Wednesday...”

Following are two slides that show you how this story would be handled by AP’s content enrichment process, before the story is distributed into various portals.

- Here is a screen shot showing the story appearing on the Yahoo Financial Site, at 8:55 p.m. Eastern Time on September 16, “The government is expected to announce...” This is based on AP’s 12th Lead Writethru of 6:49 p.m. Note the ticker symbols highlighted on the left side: AIG, JPM, and GS.

- *Honolulu Advertiser* website, showing the first Washington dateline story, which moved September 16 at 8:24 p.m. Site includes AP photo, links to related local stories, related news on the Web, and options for emailing the story.
- Here is a related, later story of 25 September on Hank Greenberg's intention to sell AIG stock as it looks in the eAP browser. The user has the ability immediately to see related photos, audio, graphics, and interactives, and much of the tagging in Workbench is unavailable here.
- This is the same story, appearing on an I-Phone, through a Web feed on the Mobile News Network.

As I hope these images make clear, the AP is growing into a far more complicated mechanism than ever envisioned by its founders, and they had their hands full with ornery pigeons and Atlantic cables that wouldn't transmit more than 8 words per minute at the start. How does this complexity affect the work of the archivist at AP in the 21st century?

In a sense, AP's business needs and archival interests are inadvertently converging in the digital age as they never did previously. Throughout the nineteenth century and most of the twentieth, the "left overs" of the news (the dispatches, clippings files, wire copy, message traffic) were considered of secondary value and survived willy-nilly. During the era of the telegraph and teletype, librarians at AP saved what they could from the flood of incoming copy. In the end, they were overwhelmed and had to dispose of most of what they saved. In 1950, they trusted microfilm to solve their space problems and began filming wire copy and throwing out the originals. At least we have the microfilm.

Another reason for AP's slowness to recognize the intrinsic historical value of its own reporting has been the nature of journalists themselves. From the earliest days of the organization, AP's leaders have been seasoned newsmen, displaying the journalist's instinct of looking ahead, not back. For instance, when Wes Gallagher became General Manager in 1962, he had had been at AP for twenty-three years and worked in twenty-six countries. He went to Europe in 1939 and covered the German march through the Baltics, the Allied invasion of North Africa, the Normandy invasion (for which he wrote the lead story), and the Battle of the Bulge. With such experience behind him, it is surprising to learn that in 1972, Gallagher said he wasn't interested in saving the Saigon Bureau records, telling Saigon correspondent Peter Arnett that "the only files that matter to us are what appeared on the A-wire." Undeterred, Arnett and Saigon bureau chief Richard Pyle packed 136 binders of original news and service message copy into metal trunks and shipped them to the States as unaccompanied baggage. In 2005, Arnett donated them to the Corporate Archives.

Gallagher's attitude toward his organization's historic copy may seem cavalier, but as a journalist, he wrote in order to hit Page One, not so his stories could reside in a vault. Most likely, he couldn't imagine what use anyone could possibly have for thousands of pages of copy that had ended their useful life. That has changed. The same leadership that brought AP into the digital era in 2003 also established AP's first Corporate Archives. Tom Curley understands how a usable past can reaffirm the continuity of AP's core mission, even during an era of great change. He is no longer allowing bureau chiefs free rein when it comes to making decisions about their bureau's records, (which has generally resulted in the loss of valuable materials), but has entrusted that responsibility to the Archivist. Moreover, bureau records, artifacts, and special collections are seen as potential sources of historical content that AP can mine for its own purposes.

But the real challenge is the preservation of the born-digital record. By any standard, eAP will be *the* repository of AP reporting and must be preserved. AP's Director of Global Infrastructure

assures me that it will, for the simple reason that it is much easier (and thus less costly), to go on saving files than to delete them. His mantra is “organic convenience.” It is more convenient to save everything forever than to discard select files.

How big is everything? Currently, eAP contains almost 1 terabyte in text stories, and the import of the Text Archive will add another 2 terabytes, bringing the overall database to about 3 terabytes JUST FOR TEXT! And the count will grow. The management of this asset will require tremendous skill and vigilance on the part of AP’s technical teams, for as we know, 1’s and 0’s are more fragile than newsprint.

AP has always been a pioneer in news technology for the industry. It introduced carbon paper and the typewriter into wide use; it worked with the Morkrum Company of Skokie, Illinois to develop the “printing telegraph” or teletype as early as 1914. It was the first to use the news wire to transmit photographs in 1935, and without AP, the digital camera would have been much slower to the consumer market. Perhaps it is time for AP to join the effort to discover the technologies that will preserve the data it creates.

AP has already made a start in this direction by agreeing to participate in the Center for Research Libraries National Science Foundation-funded project to study best practices for long-lived digital repositories. I believe AP’s experience may be valuable as the information professions merge and seek a common approach to digital preservation. Happily, it appears that much of what AP has been doing it has been doing right. While AP has been managing its digital files as assets, not artifacts, the results are—so far—the same.

Another recent development has been the collaboration between the Corporate Archives and the Office of the General Counsel. For the first time, AP headquarters has both a legal team and

an archivist under the same roof; historically, AP used outside counsel. Both departments have been advocating a records management program, the lawyers in order to comply with existing statutes, and the archivist in order to ensure that records of enduring value be preserved.

As AP focuses more of its resources on keeping its born-digital record legible and accessible, the archivist's role will be to ensure that format is not destiny; that the paper, microfilm, and artifact are accorded the same respect (and resources) as the digital file. And we must ensure that digital files are not merely managed and "backed-up," but intentionally preserved, and that resources are set aside for the purpose. AP's data is the news of the world.

