

Valerie S. KOMOR*

* Director, AP Corporate Archives.

When is the News Old? Appraising and Preserving the Legacy of the Associated Press

KOMOR, Valerie S., *When is the News Old? Appraising and Preserving the Legacy of the Associated Press.* Atlanti, Vol. 18, Trieste 2008, pp. 277-286.

Original in English, abstract in English, Italian and Slovenian, summary in English

Why did it take nearly 160 years for the world's largest news organization to establish a Corporate Archives? What impact has the lack of an archives program had on appraising AP's historic textual record, in both paper and electronic forms? What form should appraisal take in the future? The author will examine some of the reasons why AP delayed so long in preserving its past, including the nature of the news culture, the forward-looking attitude of AP General Managers and CEOs, who have all been journalists, and the perennial space shortages faced by both AP bureaus and headquarters in New York that resulted in lost records. She will then describe the beginnings of appraisal decisions by the first news librarians in 1904 and the solutions these librarians thought they had found in microfilm. Finally, the origins of AP's Text Archive in 1992 will be outlined. A matchless digital repository of 65 million stories, the Text Archive poses an immense challenge to those who would manage it as an asset and to those who would preserve it for its enduring historical value.

Since its founding in 1846, the Associated Press has been documenting world events for American newspapers and international subscribers. But only in 2003 did the world's largest news-gathering organization establish a Corporate Archives department, creating an internal mechanism for collecting, preserving, and making available AP's own historic records. Why the long delay for an organization with such a rich past? What impact has the lack of an archives program had on appraisal? And what form will appraisal take in the future? My focus will be primarily on the print news report (in paper and electronic form), although preservation efforts now extend to all of AP's news formats - photographs, audio, film, video, and graphics. The leadership that established the Archives is actively supporting these initiatives¹. It understands how a usable past can reaffirm the continuity of AP's core mission, even during an era of great change.

That core mission, ironically, goes far to explain why AP waited 157 years to appoint an archivist and bring its history out of hiding. While many of AP's correspondents and editors see events in their historical context, their task is to explain the present factually and without bias. They do this with a combination of experience and subject expertise, from bureaus across the United States and in 97 foreign countries. Accuracy and speed, their paramount objectives, are enshrined in the By-Laws:

The Associated Press is a mutual and cooperative association formed to gather with economy and efficiency an accurate and impartial report of the news.... [T]he news... shall be as objective and complete as human endeavor can make it².

In other words, "Get it first, get it right, and get it out of town." Any reflections on cosmic significance come later.

AP's origins lie in the quest for speed. In the spring of 1846, Moses Yale Beach (1800-68), publisher of the *New York Sun*, established a pony express to carry Mexican War dispatches ahead of the Great Southern Mail between Mobile and Montgomery, Alabama³. The riders were not paid unless they gained a 24-hour edge over the mail coaches, which they mostly did. In Montgomery, the dispatches rejoined the mail coach for the 700-mile journey to the telegraph head near Richmond. Beach offered an equal share in the pony ven-

1. Kelly Tunney, who retired in 2005 as an AP Vice President, encouraged the establishment of a Corporate Archives for many years. CEO Thomas Curley made it a reality in 2003.
2. Associated Press Charter and By-Laws, 2006. (42nd Edition). AP Corporate Archives.
3. Moses Sperry Beach, autograph memorandum, June 1872. Moses Sperry Beach Papers. The AP Corporate Archives. This manuscript provides evidence that AP was founded two years earlier than AP and many journalism historians had thought. AP acquired the Beach Papers in 2005.

ture to the New York daily papers, turning newsgathering from a competitive into a cooperative activity. Four dailies accepted the offer, and soon the five papers were being called the “associated Press of this city”⁴.

Meanwhile, in Chicago, a regional cooperative grew up in 1861 known as the Western Associated Press. In 1892, that organization incorporated under Illinois law, laying the foundations for the modern AP. It opened its 1894 annual report by confessing that “management has been so pressingly occupied with the making of Associated Press history as to leave but little time for recording it”⁵. That was an accurate statement with regard to its own history, but in reality, AP had been recording history like never before. The timely combination of Samuel F. B. Morse’s telegraph and Beach’s news service meant that information that had formerly moved only as fast as the fastest horse was being sent at nearly the speed of light. Strangers living far distant from each other could acquire the same information at the same moment. Communications scholar Menahem Blondheim suggests that AP’s peculiar strengths helped foster an American national identity in the second half of the nineteenth century, writing:

[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.⁶

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 from the Albany bureau in 1939 and covered the German march through



Jeremiah Gurney (American, 1812-1886).
Moses Yale Beach and Nancy Day Beach, 1855.
Daguerrotype. Courtesy National Portrait Gallery,
Washington, D.C.

KOMOR, Valerie S., Quando invecchiano le notizie nuove? Selezione e conservazione del lascito della Associated Press. *Atlanti*, Vol. 18, Trieste 2008, pp. 277-286.

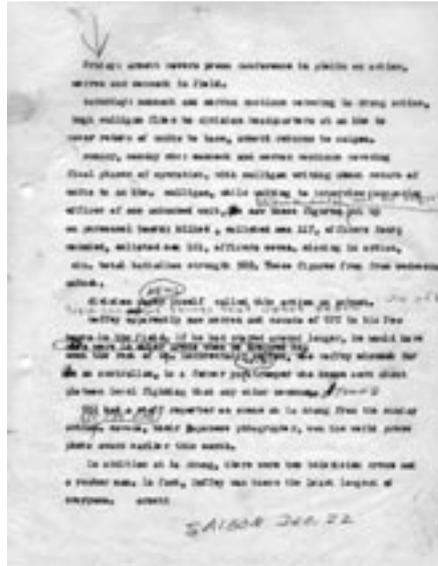
Questo articolo parlerà delle problematiche relative alla selezione ed alla conservazione in relazione alle notizie quotidianamente raccolte e trasmesse dall'Associated Press. Sin dalla suo inizio, che data al 1846, l'Associated Press ha documentato il mondo per i quotidiani americani e per gli abbonati internazionali. Al volgere del XXI secolo, l'AP è una cooperativa senza fini di lucro con 240 uffici negli Stati Uniti ed in 96 paesi esteri. Serve 1.700 quotidiani statunitensi, 850 radio notiziari e 5.000 fra radio e televisioni, di cui un decimo all'estero. Solo negli ultimi cinque anni l'AP ha iniziato a documentare le proprie operazioni. Molte sono le ragioni di questo straordinario ritardo. Innanzitutto (e soprattutto) l'AP è un'agenzia di stampa, ed il grosso delle sue risorse va nella raccolta e diffusione a trecentosessanta gradi di notizie. "Prima di tutti, meglio di tutti, e per tutti" è ancora il motto dell'AP. Nel 2003 l'AP prese la decisione fondamentale di tenere un proprio archivio, ivi compreso un sistema interno di raccolta, conservazione, fruizione dei propri documenti storici. Oltre a ciò, l'era digitale ha innalzato significativamente il profilo pubblico dell'AP, ed allo stesso tempo ha reso possibile la conservazione e la fruizione di dati prima giacenti nei magazzini. Notizie in ogni formato - testo, foto, grafica, film, e digitale - vengono classi-

4. Victor Rosewater, *History of Cooperative News-Gathering in the United States* (New York: D. Appleton and Co., 1930), 66.

5. The Associated Press, *First Annual Report of the Board of Directors to the Stockholders* (Chicago: 1894), 7. AP01, Records of the Board of Directors. AP Corporate Archives.

6. Menahem Blondheim, *News Over the Wires: The Telegraph and the Flow of Public Information in America, 1844-1897* (Cambridge, MA: Harvard University Press, 1994), 195. Blondheim was the first scholar to consult the Moses Sperry Beach Papers in private hands and conclude that they document the origins of AP in 1846 as a direct response to the telegraph and the popular desire for Mexican War news.

ficcate da un pionieristico team di bibliotecari per facilitare il recupero istantaneo da un database centrale. In un mondo “insaziabile di informazioni”, come il presidente dell’AP e del CEO Tom Curley lo ha definito, l’abilità di selezionare e personalizzare le notizie determina ora il modo in cui le notizie vengono trasmesse, e, alla fine, come vengono conservate. Questa è la più grande rivoluzione nel campo dell’informazione da quando il telegrafo ha iniziato a diffondere notizie alla velocità della luce, sorpassando il pony e la ferrovia in un balzo. La risposta alla domanda “Quando invecchiano le notizie nuove?” guida gli archivisti dell’AP nel selezionare le notizie per la conservazione permanente. Come disse una volta un cinico, “Le notizie di oggi sono i cartocci di domani”. Questo non è più vero. Sebbene le notizie diventino “vecchie” praticamente nel momento stesso in cui vengono diffuse, esse raggiungono velocemente lo status di “eredità”, una risorsa preziosa per gli storici. Fornirò una definizione di “notizia” nel contesto del “circuito dell’informazione”, tracciando il diagramma del suo ciclo vitale. Spiegherò inoltre come le misure in vigore conservino le notizie attraverso vari formati e piattaforme (nel formato originale, in microfilm, nei giornali, nelle agenzie, come Newsbank o ProQuest, e nell’archivio testuale della stessa AP). Chiuderò con un breve sguardo alle sfide affrontate dall’Archivio Corporativo dell’AP durante la ricerca di questo patrimonio disperso, mentre inizia la collaborazione con AP, al fine di incoraggiare la conservazione e la fruizione. L’AP non si vede come un tesoro nazionale o un deposito cultu-



Peter Arnett, AP war correspondent. Memorandum on coverage of the battle at Ia Drang, 22 December 1965. Saigon Bureau Report. AP Corporate Archives.

the Baltics, the Allied invasion of North Africa, the Normandy invasion (for which he wrote the lead story), and the Battle of the Bulge. After the war, he became Paris bureau chief and supervised the establishment of the German service. With such experience behind him, it comes as a surprise to learn that in 1972, Gallagher said he wasn’t interested in saving the Saigon Bureau records, telling top correspondent Peter Arnett that “the only files that matter to us are what appeared on the A-wire”⁷. 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 Arnett’s unaccompanied baggage. In 2005, Arnett donated them to the Corporate Archives. The files are of incalculable value, as they provide a 12-year chronological record of the Vietnam War, the most comprehensive day-to-day account of the conflict that exists anywhere. 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. And he was not alone in his views. Every General Manager and CEO who followed him, until the arrival of Thomas Curley (from outside AP) in 2003, shared it. AP was a news agency, not a museum, went the thinking. The concept of an “archives” probably didn’t enter their minds.

Arnett’s rescue of the Saigon files is just one instance of an individual acting independently to preserve AP’s past. In general, a flurry of collecting occurred at obvious historical moments, when staffers would realize that what was coming off the teletype *was* history. In this way, Joe Pickle, Managing Editor of the *Big Spring Herald*, saved 14 sheets of wire copy reporting the surrender of Germany on May 7, 1945. Sixty years later, his son gave them to the AP. An uncut roll of the broadcast wire copy for November 22, 1963 was saved in a shoebox by a North Dakota radio station employee. After forty years, that made its way to New York. All of the “A” or national wire copy from the day President John F. Kennedy was assassinated bears the type-written command, “Preserve,” and it was, throughout the AP and by its member newspapers.

The bureaus have had practical and legal reasons for saving their records. News copy and reference files are vital sources in the preparation of stories, and bureaus must comply with state statutes

7. Peter Arnett, Letter of transfer, December 18, 2005. AP Corporate Archives.

of limitations on bringing libel or defamation suits⁸. Large troves of copy accumulated in Albany, Cape Canaveral, Chicago, Miami, Pittsburgh, and Washington, D.C. Washington has always cultivated its own identity as the largest bureau outside New York headquarters and as the nerve center of national political reporting. This may explain its decision in 1944 to donate its text report from 1915 to 1930 to the Library of Congress. The material consists of 375 bound volumes of hand-edited carbons (text before it moved on the wire), and includes the full gamut of stories, from weather and agricultural reports to the sinking of the liner *Lusitania*; passage by the House of the war resolution at 3:00 a.m. on April 6, 1917; and the outbreak of the Bolshevik Revolution in Russia⁹.

Washington probably transferred its files due to lack of space, a frequent predicament for AP bureaus, many of which relied on member newspapers for office space. This meant that bureaus set up shop in the mornings in the offices of the afternoon paper and in the evenings packed up their news materials and took them to the morning paper down the street. Needless to say, no one had room for overflow copy.

Even headquarters felt the same pressures. Between 1849 and 1925, AP had at least six addresses in Manhattan, and each move meant the loss of records. And yet, space shortages determined appraisal decisions both before and *after* AP's move to Rockefeller Plaza in 1938. In a candid admission by Board Secretary Lloyd Stratton in 1945, we learn that

the bulk of old files had reached impossible dimensions, going back as [they] did to the early days of the Illinois Corporation in 1893. To secure needed storage room, segregation had been made in the files covering the first twenty years, namely the files of the Illinois Corporation 1893-1900 and of the New York corporation from 1900-1913. From these early files all the corporate records had been segregated. The Board approved the preservation of the segregated corporate records, and directed that the balance of the old files be destroyed¹⁰.

We can only guess what the "balance of the old files" would have told us. But at least the loss was not total.

Oddly enough, benign neglect led to the preservation of the Archives' core. Four hundred linear feet of the AP General Files were put out of sight and eventually out of mind in the late 1970s, when they were sent to the sub-mezzanine basement of 50 Rockefeller Plaza. They were discovered just after the appointment of AP's first archivist, on the eve of another headquarters move in 2004. The General Files, so-called because they were created by AP's General Office, are filled with editorial, administrative, and bureau correspondence from 1920 to 1970, and include those early corporate records that had been spared destruction in 1945. Wire copy is particularly prominent in these files. If an editor had some concern about an AP story, or a member had written to complain about a story, the copy got saved. Thus, we have fairly complete files relating to the integration crisis of 1957-58 throughout the South, since Executive Editor Alan Gould was in almost daily correspondence with

rale, ma nel sostenere la conservazione del proprio patrimonio sta divenendo più di una agenzia di stampa. Sta diventando un modello per la comunità archivistica per assicurare accesso continuo alle notizie nel momento in cui esse diventano storia.

KOMOR, Valerie S., *Kdaj novica zastara? Vrednotenje in hramba gradiva novinarske agencije AP.* Atlanti, Zv. 18, Trst 2008, str. 277-286.

Zakaj je trajalo skoraj sto let, da so osnovali enoten in skupen arhiv? Kaj bo vplivalo na dokumente, ki bodo za zgodovino pomembni iz obeh področij: klasičnega in elektronskega? Kakšna naj bo oblika vrednotenja v prihodnosti? V prispevku avtor prikaže nekaj vzrokov, zakaj se je novinarska agencija AP šele po dolgem času odločila, da obrani dokumentacijo iz preteklosti, po drugi strani pa da zavaruje naravo kulture novic in novinarstva in da gleda tudi v bodočnost novinarskega poklica na podlagi vrednotenja preteklosti. Avtorica zato opisuje začetke vrednotenja prvi novinarskih vesti iz leta 1904, ko so prenašali dokumentacijo na mikrofilm in zaključuje s prikazom originalov tekstovnega gradiva iz leta 1992 in s prikazom neverjetnega digitalnega skladišča z 65 milijoni zadetkov, kar predstavlja neizmerno izziv za vsakogar, ki hoče upravljati s tako veliko dokumentacijo.

SUMMARY

This paper will address issues of appraisal and preservation as they relate to the news gathered and transmitted daily by the Associated Press. Since its inception in 1846, the Associated Press has been documenting the world for American newspapers and international subscribers. At the turn of the 21st century, AP is a unique not-for-profit cooperative, with 240 bureaus in the United States and 96 foreign countries. It serves 1,700 U.S.-based newspapers, 850 radio news affiliates and 5,000 radio and TV outlets, a tenth of them abroad. Only in the last five years has AP begun to document its own operations. There are several reasons for this extraordinary delay. First and foremost, AP is a news agency, and the bulk of its resources go to the 'round the clock gathering and transmission of news stories. "Get it first, Get it right, and Get it out of

8. Under New York State Civil Procedure Law and Rules, a cause of action for libel and slander must be commenced within one year of publication. The statute of limitations varies from state to state, from one to three years, and therefore impacts differently on AP domestic bureaus, which have to decide how long to keep materials used in preparing stories. To date, AP's advice to each state has been to use the local statute of limitations, plus three months, as the benchmark.

9. Associated Press Dispatches, 1915-1930. MSS5066. Manuscript Division, Library of Congress, Washington, D.C.

10. Board Secretary Lloyd Stratton, *Abstract of Miscellaneous Board Resolutions*. AP01, Records of the Board of Directors. AP Corporate Archives.

town," is still the AP mantra. In 2003, AP took the pivotal step of establishing its first Corporate Archives, thereby embracing an internal mechanism for collecting, preserving, and making available AP's historic records. In addition, the digital era has raised AP's public profile significantly, and at the same time made it possible to preserve and make accessible information that previously languished in basement filing cabinets. News in all formats - text, photos, graphics, film, and online video - is being classified by a pioneering team of librarians to facilitate instant retrieval from a central database. In a world "insatiable for information," as AP President and CEO Tom Curley describes it, the ability to select and personalize news now determines how news is transmitted, and, ultimately, how it will be preserved. This is the greatest revolution in information since the telegraph first moved news at the speed of light, surpassing the pony and the railroad in one bound. The answer to the question, "When is the news old?" guides AP archivists as they appraise the news for permanent preservation. As a cynic once said, "Today's news is tomorrow's fish-wrapper." That is not true any longer. Although the news becomes "old" almost as soon as it moves on the wire, it quickly attains the status of "legacy" or "archival" content," a valuable resource for historians. I will provide a definition of "news" in the "wire service" context, and diagram its life-cycle. I will also explain how measures already in place are preserving the news across various formats and platforms (in original wire copy, on microfilm, in newspapers, in subscription-based news aggregators, such as Newsbank and ProQuest, and in AP's own Text Archive.) I will close with a brief overview of the challenges faced by the AP Corporate Archives as it seeks physical and intellectual control over this dispersed content, while initiating partnerships across AP to encourage preservation and access. AP does not see itself as a national treasure or a cultural repository, but in supporting the preservation of its legacy, it is becoming more than a news agency. It is becoming a model for the archival community for ensuring persistent access to the news as it becomes history.

southern publishers who found fault with AP's reporting. In this roundabout manner, some remarkably historic wires have been preserved. Doug Cornell's copy giving the back and forth oral re-argument in *Brown v. Board of Education* in December 1953 survives, as does Relman Morin's dramatic Pulitzer Prize-winning dispatch from Little Rock of September 25, 1957.

The Appraisal of AP's Textual Output

As closely as we can determine, formal "appraisal" decisions began at AP in 1904 with the establishment of a news reference library that would serve the news department. Before describing what the library collected, I will describe the types of paper documents that AP has created in the process of writing and transmitting the news. They are: the dispatch, the edited copy before it moves on the wire, and the wire copy.

AP provides 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. For most of its history, AP covered the news in cycles, for morning and afternoon newspapers. Stories developed during the morning for afternoon papers, and were then rewritten for the next day in versions called "night leads." Stories originated with the reporters' dispatch or draft, were edited for the wires and filed. Updates during a news cycle were (and are) called "leads." When high-speed delivery (1200 words per minute) replaced the teletype report (66 words per minute) in 1985, updated versions were replaced by "writethrus," which eliminated the need for papers to cut and paste leads, "adds," and inserts, and news cycle versions were replaced by a "BC" or "both cycles" version.

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. Wire copy is just that: copy that flows over the wire via teletype to the newspaper, and simultaneously back into the AP offices, providing a record of what was sent. At AP, there have historically 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 state wires in every state. 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, even into the digital age.

The first dispatches were written by hand and are now quite scarce. A May 19, 1864 letter to the editor of the *New York Times* describes AP dispatches as "written in a peculiar style, being manifolded on tissue paper"¹¹. The earliest dispatch I have come across is one datelined New Orleans, January 20, 1882; it resides in the William Henry Smith Papers at the Indiana Historical Society in Indianapolis. Smith was the General Manager of the Western Associated Press from 1883 to 1893. The ink is iron gall on a machine-made laid

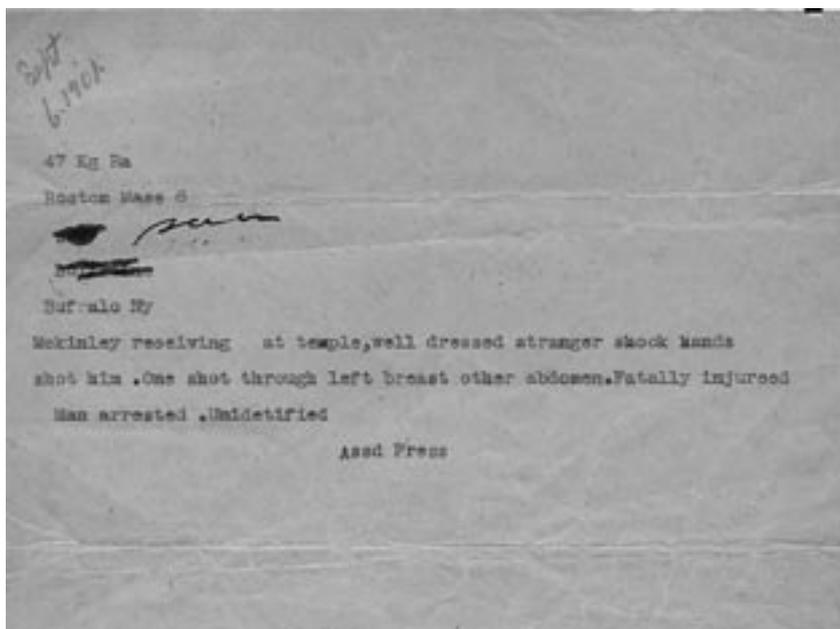
11. William Cowper Prime, David M. Stone, David Hale, Gerard Hallock, and Manton Marble, to the editors of the *New York Times*, May 19, 1864. *New York Times* Online Archives.

paper; the careful hand suggests a final draft.

From 1846 until 1914, dispatches were transmitted by telegraph using Morse Code, on flimsy wood pulp paper. Sending operators took written dispatches and translated them into the dash-dot language of code. Telegraph keys sent signals at a rate of twenty-five to thirty-five words a minute. In newspaper offices, receiving operators rendered the code symbols back into words and set type accordingly. The Archives holds an AP dispatch sent by telegraph from Boston to New York on September 6, 1901, reporting the attempted assassination of President McKinley. The ink scrawl seems to read "Save".

The Russo-Japan war of 1904-05 may have spurred the first deliberate documentation of AP's war coverage and the creation of a news library to organize it. That war was the first major conflict after AP incorporated in New York State, and AP General Manager Melville Stone had committed significant resources to covering it. Early in 1904, he traveled to St. Petersburg, where a new bureau had been established, in order to secure from the Czar favorable cable rates for AP dispatches, precedence in sending AP telegrams, and access to all Russian government entities for the purpose of obtaining the news¹².

Associated Press Dispatch, January 20, 1882. MO258. William Henry Smith Papers. Indiana Historical Society, Indianapolis, Indiana.



AP Cable reporting the shooting of President William McKinley, September 6, 1901. AP Corporate Archives.

AP's first news librarians assembled a large clippings collection which they organized by subject and indexed numerically on catalog cards. The earliest scrapbook, compiled by the Burrelle's Clipping Bureau, dates to 1898 and bears the seal of a Chicago stationer. After moving to New York in 1900, AP continued to use the bureau to clip both AP stories and stories about AP. These immensely detailed articles describe the news technology of the first decades of the twentieth century, as typewriters entered the newsroom, the cyclostyle duplicator made copying telegraph dispatches possi-

12. *Associated Press Annual Report*, 1904. (New York: Associated Press), 3.

ble, and cable, leased wires, and wireless joined with telephones, pneumatic tubes, airplanes, and messengers to bring news into New York from around the world.

AP Begins to Save its Wire Copy

The library continued to save and index clippings through World War I, well after the teletype was introduced¹³. With the teletype, however, AP had a steady stream of its own words flooding into headquarters, bureaus, and members' newsrooms. As the copy rolled off the incoming printers, copy boys cut and delivered it to the editing desks. Once transmitted, wire copy was "spiked" to be retrieved at the end of each day and bound together with a brass brad. These bulky bundles quickly filled file cabinets to overflowing, and most were eventually disposed of. The news librarians made their selections from this daily influx, pasted them onto boards and indexed them, just as they had done with clippings. In 1920, AP's *Service Bulletin* featured a profile of Harold Martin, Superintendent of the Eastern Division, whose office was "equipped with a steel filing cabinet containing complete duplicate copies of every item of news received by and sent out of the New York office for every day in each month and so arranged as to be readily accessible for instant reference"¹⁴. No doubt other supervisors had a similar arrangement.

Sports Desk Wire Copy ("paste-up"), 1974. The Sports Desk retains its own wire copy from 1964 to 1988.



The first indication that librarians had been disposing of wire copy comes in 1948, when news librarian Wayne Cottingham disclosed that "everything connected with World War II will be preserved indefinitely for reference use. There will be no more of the ten-year discard business that took away all the valuable story of World War I and what happened in its wake"¹⁵. Cottingham thought he had found a way to reverse these admittedly unfortunate losses.

"Microfilm saves considerable space," Cottingham explained. "For example, contents of nearly two drawers of a legal-size filing cabinet go on a roll of film small enough to hold in your hand"¹⁶. And you can save even more space by discarding the originals after filming, which the librarians did.

In the late 1970s, AP contracted with UMI to microfilm stories dating between 1937 and 1974. It later made a license agreement with UMI to market the material to libraries as "a complete history of the past 30 years, as reported by Associated Press correspondents around the world"¹⁷. Today, the microfilm collection comprises over 900 reels. In addition to the A wire, there are discrete sets of film for AP NewsFeatures, New York Bureau stories, and Sports stories. The card indices continues to provide access to the microfilm on the very rare occasion that it is used by the archivists who are now *de facto*

13. *Ibidem*.

14. *The Service Bulletin* (New York: Associated Press), vol. 54, 1920, 18.

15. Wayne Cottingham, *Microfilm Magic Saves Old Files*, AP World, Spring, 1948, 32. AP Corporate Archives.

16. *Ibidem*.

17. From the author's office files.

custodians of AP's physical text archives. Though incomplete, the film remains the only extant version of AP's "A" wire copy for half of the twentieth century¹⁸.

The Text in Electronic Form

A mere ten years after its agreement to let UMI market its microfilm, AP began to formulate plans for an electronic Text Archive¹⁹. News Technology and editorial staff had dreamed of a central digital repository that would allow quick access to the national, state, and international reports from anywhere in the world. AP wanted a storage system that reporters could access as they were composing their stories. Legal concerns were also a stimulus for the project, as AP needed - as it always had - to point to its copy as proof it had filed a story. Work began on the project after the general election of 1992 and took three years of development²⁰.

To build the Text Archive, three sources of existing electronic text were migrated into one repository: "permanent" databases (1996--); legal and research "Felix" databases (1995); and VuText databases (1985-94). Developed and owned by Knight-Ridder, VuText was AP's first foray into digital text when it began using it in 1979. AP reporters simply typed their stories into the newsroom editing system, and a copy went into VuText²¹. By 1996, VuText included stories from the national, financial, New York Bureau, and sports wires for the period 1985 to 1994; it did not include any pre-1985 content. It took some years to add the complete VuText database to the Text Archive, but by 1995, the national and state reports were available. In 1997, version histories for each story were added, as was AP Broadcast copy (copy written for AP radio and television members). International wire copy in French, Spanish, and German was added in 1998, and Dutch in 1999.

On February 13, 1996, the Text Archive went live at AP, and it is still in place. At about the same time, an editing tool known as Reporter's Workbench was rolled out. Workbench, an interface client program, resides on the reporter's PC and accesses the news production system and the Text Archive at the same time. When an editor sends a story to the wire, it automatically goes into the Text Archive, together with any messages, story lists, and advisories not intended for publication. This kind of internal message traffic adds valuable context to the news, and paper versions exist throughout the Archives. By any standard, then, the Text Archive is a matchless repository of AP reporting and internal communications and must



Screen Shot, Reporter's Workbench editing tool, 2008.

18. In 1977, AP began selling its "A" wire report (just final stories) to Lexis Nexis.

19. Telephone interview on June 3, 2008, with Bruce Toll, part of the team that developed software to run the Text Archive.

20. Telephone interview with Tim Gallivan, retired AP Systems Editor, May 9, 2008.

21. Vicky McCargar, *Digital Preservation Consultant*, in an unpublished report for an ongoing Center for Research Libraries study launched in 2007 on long-lived databases, 30.

be preserved permanently. And in the world of electronic databases, I am assured, it will be, for the simple reason that it is much easier (and thus less costly), to go on saving files than to delete them. The new mantra is “organic convenience”²². It is more convenient to save everything forever than to discard select files.

But we have to ask, what is *everything* and how big is it? The Text Archive itself holds 65 million stories²³. This is an estimate of the number of text documents currently in the Archive, including both domestic and world (foreign language) databases. A document is a unique story version, so each writethru and lead is counted separately. As more documents are added to the Archive, the count will grow.

As the Text Archive grows, another major technology initiative is underway that will require the migration of the Archive into a single search-and-delivery database. Electronic AP, or eAP, will categorize, search, and distribute AP news content across all formats: text, photos, audio, video, and graphics²⁴. A team of information specialists is developing a 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 other subjects. In a world “insatiable for information,” as AP President and CEO Tom Curley describes it, the ability to select and personalize news will determine how news is transmitted. But it will also determine how it will be preserved.

Current plans call for the Text Archive to be migrated in its entirety into eAP and run through the eAP classification services. Digests, advisories, and writethrus will all be migrated, so that eAP will conserve the versioning structure of AP stories, with each story version transformed into a single APPL document. The linkage between the different versions of a story will be maintained.

It is important to note that although AP is becoming a news repository on a grand scale (on any given day, its 3,000 servers receive more than 100 terabytes of data), it does not see itself that way. It is a business, and its decisions are made in light of business and legal requirements. However, even if eAP had not been developed, AP would still need to move the Text Archive to a new technology platform to combat the obsolescence of the existing hardware and the increasing difficulty of supporting, enhancing, and scaling the current software platform. AP would take these steps not primarily to *preserve* its Text Archive, but because the Archive is an asset that must be managed indefinitely to AP’s benefit and that of its members. In a real sense, AP business needs and archival interests are inadvertently converging in the digital age as they never did previously. Throughout the 19th and the first 85 years of the 20th century, the “left overs” of the news (the dispatches, clippings files, wire copy, message traffic) were considered, in the end, junk. Now all of this material is seen as a manageable asset. What *kind* of an asset remains to be seen; certainly monetizing text will never be a core business for AP.

22. The words of Dan Raju, Director of Global Infrastructure for AP, and in charge of finding storage for AP’s burgeoning digital output.

23. Mike Alexander, *AP Enterprise Application Services*, June 9, 2008.

24. AP Press Release, June 28, 2004. http://www.ap.org/pages/about/pressreleases/pr_062804b.html

But one thing is clear: AP's digital text *is already* a valuable archival resource.

Just how institutions like AP are managing large amounts of data is a question that is currently receiving serious study. The Center for Research Libraries has just initiated a two-year National Science Foundation investigation to analyze eight established "long-lived" collections of data, and create tools for developing and assessing new repositories²⁵. AP is participating, as are the American Chemical Society, UMI ProQuest Dissertations, the General Social Survey, the U.S. Geological Survey, Sloan Digital Sky Survey, the National Center for Atmospheric Research: Earth Observing Laboratories, and the Arabidopsis Information Resource. The very fact of AP's involvement in such an effort is a sign of changing times, suggesting that AP sees itself well-placed to contribute to, and benefit from, research into digital preservation. It is surely a sign that AP recognizes its early and longstanding influence in pioneering communications technology, from ponies to web portals, and in bringing those technologies to its members.

As AP focuses its resources on keeping its born-digital record legible and accessible, it is well to recall that it did not always keep the same commitment to its paper and microfilm. The reasons for this are not simple, as we have seen, but the future role of the archivist at AP will be not be simple either. We need to ensure that both analog and digital formats are equally respected, and that the digital record is not merely managed but carefully preserved for its enduring historical value²⁶.

25. <http://www.crl.edu/content.asp?l1=13&l2=58&l3=174>. See also http://www.nytimes.com/2008/04/09/technology/techspecial/09store.html?_r=2&ref=techspecial&oref=slogin&coref=slogin.

26. Katie Hafner, *History, Digitized (and Abridged)*. *The New York Times*, March 11, 2007, Sunday Business Section, 1. Hafner's article takes on the illusion that all information is on the Web. In fact, there are vast amounts of materials in archives and libraries that will never be digitized, due to money, technology, and copyright complications, among others.