A mandate to preserve

Assessing the inaugural Newspaper Archive Summit

by Victoria McCargar

Rescuing orphaned and DIGITAL content

SPONSORS

rji donald w. reynolds journalism institute Missouri School of Journalism University of Missouri

Mizzou Advantage University of Missouri

LIBRARY OF CONGRESS

MU Libraries University of Missouri

Table of contents

Table of contents 1

Executive summary 4

1. Impetus: Why a 'rescue' theme? 6

Preserving newspapers 6

Preserve what? 6

Who cares? 7

Thirty years of technology 7

Fragile formats 8

2. Stakeholders and 'preservation' 9

Two disciplines 9

Historians and genealogists 9

librarians and archivists 10

Newspaper archivists 10

Publishers 12

Vendors 12

Stakeholders and the Summit 12

3. The digital challenges 14

Issues in news preservation 14

Urgency 15

Microfilm: going, going… 15

4. The Summit: agenda 16

Agenda 16

Themes 16

Stakeholders 16

Funding, funding, funding 17

Brainstorming 17

5. Summit outcomes 19

Searching for commonality 19
What is 'preservation' 19
Are the right people in the room? 20
Reactions to panels 20
Sessions 21
6. Post-summit results 23
Coverage of the Summit 24
7. Looking ahead: Summit II 25
Rescue again: What does it mean? 25
Creating focus 25
Archives of defunct newspapers 25
Online only 26
Why should we care? 26
Seeking common interests 26
Topics 27
Reactions from Summit I 27
Topics for Summit II 28
The role of recruitment and RJI 29
Conclusion: Building on Summit I 30
Next time? 30
Social networking 30
Educating stakeholders 31
Advancing the Research 32
The end – for now 32
Appendix A: Agenda 33
Appendix B: Brainstorm groups 35
What is in a newspaper archive? 35
What is in a newspaper photo archive? 37
IDEA GENERATION 38

IDEA DEVELOPMENT 40
Appendix C: Surveys 44
Reynolds Journalism Institute Survey - June 24, 2011 47
Newspaper Archive Summit Impact Survey Results 47
About the author 54

Executive summary

Historically, when a newspaper ceased publication, the photographs, clippings and bound volumes were handed off to the local historical society or public library. They sat there, and many continue to sit there, until the organization decided what to do with them.

While newspapers have been around for centuries, they weren't considered worthy of preservation and indexing until the late 19th Century. Today, newspapers are largely digital. Their content is multimedia, and there is a dwindling presence of the physical edition as news moves online. Unfortunately, this revolution in newsgathering has made obsolete the tried-and-true archives methods (known as benign neglect), and along with them the old handoff paradigm. Instead, historical societies and libraries are struggling to deal with this new digital content and how it will be preserved for future generations. The fact that digital archives are much more fragile than paper ones is a problem of which many publishers are completely unaware.

On April 10-12, 2011, the Donald W. Reynolds Journalism Institute at the University of Missouri, MU Libraries and Mizzou Advantage, and the Library of Congress, brought together 125 professionals from archives, historical societies, libraries – and, yes – journalism, to discuss the end of benign neglect and the growing need to rescue threatened news collections. This was the first time, to my knowledge, that the Fourth Estate has been so represented, and at a high level, and it was a tremendous achievement just getting publishers in the room.

This paper reviews the issues of newspaper preservation confronting us today and how they were applied in the structure and content of the Newspaper Archive Summit held in Columbia, Missouri. Outcomes and takeaways of Summit I are discussed, as well as the potential shape and content of Summit II.

Among the takeaways of Summit I were:

- Stakeholders from diverse professions (historians and genealogists, librarians, journalists, news aggregators) have **confusing and often conflicting ideas about "digitization" and "preservation."** This points to a need for basic education with the goal of developing unified interest around key topics and priorities.
- There is a need to **identify and prioritize risks of threatened news collections**, both publicly and privately owned.
- Given the pace of downsizing and number of papers walking away from print publication, **time is of the essence.** The longer digital collections sit in storage, the greater the risk they'll be lost altogether.

- A large body of research worldwide has gone into digitizing newspapers printed before 1923. It's time to start applying it to the current situation.
- Necessary key components of basic education include **economics of ownership, copyright and the cost of preserving (or not preserving).**
- Recruitment for Summit II is of prime importance to ensure that the right voices are heard – and are listening.
- The journalists at Missouri and other schools will need to play an even larger role and reach out to their peers at newspapers. **The discussion must be elevated above "it's a library problem." It's also a publishing problem, a journalism problem.**

The mission of the Newspaper Archive Summit was to get the conversation going. In that, it certainly succeeded. The presence of so many competing demands and what some referred to as "biases" merely illustrates that preservation in general (and of newspapers in particular) is a very wide discussion indeed. It has many stakeholders occupying silos. But they have one thing in common: All recognize that this stuff, these news archives, this "history on the wing," this "first draft of history," is worth the difficult and urgent tasks ahead.

Vicky McCargar, Missouri '77
Encino, CA
October 10, 2011

1. Impetus: Why a 'rescue' theme?

> ***'To me the biggest achievement of this symposium was bringing these parties into the discussion.'***

Preserving newspapers

Forty years ago, changing reader tastes, commuting habits and demographics signaled the beginning of the end for the p.m.-daily *Herald-Examiner*, which had only recently fell from its position as the largest-circulation daily in Los Angeles. When the end finally came in 1989, most of its vast "morgue" went to the Los Angeles Public Library and the University of Southern California.

In those days a newly defunct newspaper could deposit its clippings and photo prints on the loading dock of the local historical society. Nowadays, the recently defunct newspaper is likelier to hand off boxes of CD-ROMs or a database. Does the willing recipient have any idea what it is getting into?

It is not insignificant that the *Herald-Examiner*'s paper archives also sat around more or less untouched for twenty years until the libraries started digitizing the collection. The problem we face is that today's digital media do not have the luxury of that kind of benign neglect.

The grinding industry recession continues to take a toll on newspapers of all sizes. Some fold up shop and disappear, others start publishing exclusively on the Internet, and the largest papers in the country started cutting staff 15 years ago and haven't stopped.

Finally, recent copyright rulings have put a lot of newspapers' digital content in ownership limbo.

All of these – changing reader tastes, economics, staff reductions and technological change – are combining to put newspaper archives at serious risk. What is threatened is literally the future of the past: What will historians, genealogists and journalists be looking at when they seek historical information a generation or two down the road?

PRESERVE WHAT?

The preservation of news is a problem that even the biggest national libraries are struggling with, from Amsterdam to Singapore. Delivery of news on the web, via Twitter, by aggregators (legitimate ones and data thieves) challenges even the definition of *news*; let alone what it means to preserve it. At the same time, massive public projects to digitize historic newspapers have been wildly successful with the public, who have come to expect that *all* newspapers will be freely available online. Meeting that kind of demand and that level of disconnect is a growing problem for under-funded institutions.

It is into this challenging arena that the Reynolds Journalism Institute stepped in with the inaugural News Archives Summit: Rescuing Orphaned and Digital Content, held in Columbia, Mo., April 10-12, 2011.

RJI and the University of Missouri are well positioned to take up the issue of news preservation. Home to the renowned Missouri School of Journalism, Columbia is also home to the State Historical Society of Missouri and Missouri Press Association, two organizations vitally concerned with the future of newspaper archives. As a major university, the University of Missouri has many of the resources necessary for preservation under one roof, from computer engineering to library science to journalism. In theory, anyway, a new repository for digital newspapers (that is, a centralized facility dedicated to preservation) would seem to be well situated.

WHO CARES?

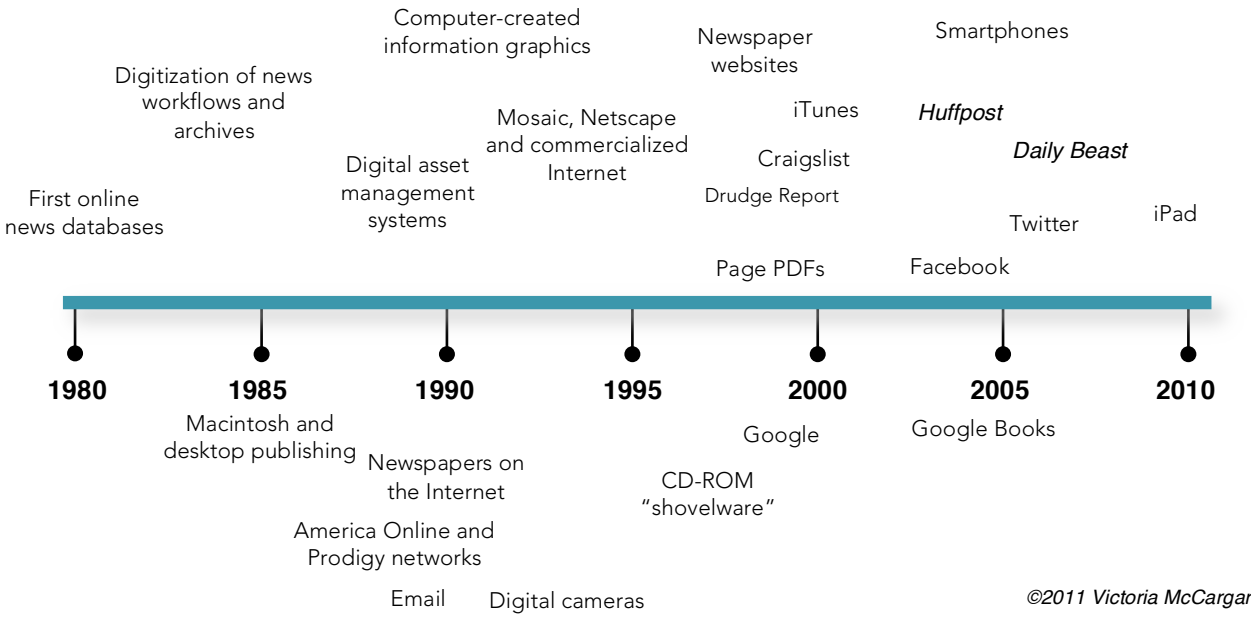
Newspapers, assembled in haste and printed on bad paper, were never intended to last more than a few days – after all, there will be another one tomorrow morning or next week. To date, worrying about preserving them has been limited to a few people in the academic library and archives sector, and news librarians. For journalists and publishers, the general threats from short-lived digital media are an interesting story they sometimes write about (such as lost NASA data), but it has never occurred to them to apply these threats to their own material.

Obviously publishers of newspapers have other things to think about – like survival. But in the numerous angst-ridden analyses of the future of newspapers, reference to making money off the archives is often a given, an assumption that seems to presuppose that the archives will be there in the first place. That assumption is not rooted in reality. The threats are very real, as we'll see.

Thirty years of technology

Newspapers today are a combination of revolutionary technology and 19th century production. Not unreasonably, the blistering pace of change has caused havoc – especially economic – in a tradition-bound industry. New desktop tools and automation processes provided not only cost savings for publishers but also created a wealth of new types of content. The same market-driven innovations that made these possible also puts a lot of material at risk, as this timeline illustrates:

Figure 1. New technologies and delivery channels



FRAGILE FORMATS

How do disruptive technologies put content at risk?

It is a process that has been going on throughout written history. The book, for example, put an end to the scroll, and any information/knowledge/wisdom that didn't make it to the next format was usually lost. Since we're no longer burying carved stone in sand, modern formats are especially short-lived – newsprint itself being a prime example of how new speed and convenience (in paper manufacture) greatly shortened life of the medium. Digital formats are even more prone to this phenomenon, and startlingly quickly. The speed and convenience of digital delivery are leaving paper and film behind, and yet there are still no reliable ways to ensure the longevity of these new platforms.

Out of the speed and convenience of digital production process have come ideas like "repurposing," "data mining" and "curating" of the newly digitized collections. Publishers and the beneficiaries of preservation, however, have often different ideas about the value of the collections. We know anecdotally that a news story's useful life is only about 90 days; after that it's seldom accessed. There is a good reason that newspaper archives were known for decades as *morgues*. The fact is, the great keepers of news archives – the newspapers themselves – do not value their assets in the same way as librarians, historians, genealogists, and others with an interest in the past.

2. Stakeholders and 'preservation'

Both the theory and the reality of preservation, especially digital, emphasize who is in the room: The continuum of stakeholders who have roles in production, management, sustainability and use of archives. They're all necessary to any discussion of preservation, but they differ in interesting ways on what the term actually means.

Two disciplines

There are basically two distinct disciplines that might both be called "newspaper preservation." One is the traditional preserver, the so-called cultural heritage institutions: Libraries, archives and historical societies. They have a long track record with conservation of embrittled newsprint and microfilming dating back to the 1930s. They are the source of the vast majority of scholarly literature in the field. Since the late 1980s, the preservation focus has been on newspaper digitization projects (from paper or microfilm), a topic that tends to dominate both professional conferences and the current literature. Scanning technology has improved immensely, and user expectations are fueling intense demand for more and more digitized information.

The other is the emerging field of digital preservation, the theory and practice of sustaining digital objects that have no analog original, also known as "born digital" archives. While segments of the cultural heritage community have been focusing for many years on the particular fragility of digital materials, the application to born-digital newspapers is in its infancy. Many librarians and archivists long accustomed to working in newspaper preservation have only minimal understanding of the life cycle of digital news and what is in store for those who would attempt to preserve it – for example, the wholesale migration of publishing to the internet, and blurring concepts like "edition."

Both groups were represented at the Newspaper Archives Summit, but the degree of overlap – or common ground – between them isn't clear. In section 7, we'll treat some issues of commonality.

HISTORIANS AND GENEALOGISTS

We can begin our review of Summit stakeholders with two of the primary beneficiaries of preservation – historians and genealogists, for whom newspaper backfiles are often the first step in research. As end users, they're counting on an infrastructure of production and preservation actors and actions to deliver what they need.

As a group, they're quite accustomed to working with crumbling clippings and scrapbooks, brittle bound volumes and poor-quality microfilm. For them, digitization of historic newspapers, photographs, church

records and other materials has been a tremendous boon. These stakeholders are most likely to understand "preservation" in terms of digitization; an old newspaper will be said to be "digitally preserved" when it has been scanned.

There is a basic omission in that understanding. In fact, with digitization, one preservation problem has actually been traded for a new one, sustaining the new digital collection. This is one of the things that make cross-domain groups so interesting. But these groups also have responsibility for helping themselves and each other shape the way digital news is delivered in the future in the face of ever-changing options.

LIBRARIANS AND ARCHIVISTS

Given newspapers' bulk and tendency to disintegrate, efforts to preserve them have been going on for nearly a century, primarily by libraries and archives with a long-range definition of value. The format of choice, of course, was and continues to be microfilm – another example of an old and probably transitional production technique combining with digital technology.

There are a couple of issues in this domain.

- *Microfilm.* The majority of the newspaper preservation community – academic libraries and state historical societies – still think in terms of microfilm, with all of its attendant downsides (not the least of which is that users hate it). For these professionals, also struggling with crumbling clippings, massive bound volumes and poor-quality microfilm, "digitization" is a much more desirable form of "preservation." They're also dealing with mounting demand for more digitization.
- *Paper-centric.* Most of the research in newspaper preservation, until quite recently, has been devoted to these ongoing problems in handling old newsprint. The massive scanning projects of the last decade – Library of Congress' National Digital Newspaper Program and its manifestations at the state level – have produced a large body of literature about scanning newspapers up to the copyright wall in 1923. It is only very recently that the problem of born-digital newspapers has been recognized in the library world, and many preservation practitioners are still uninformed about the complexities of digital collections.
- *Vendor roles.* Libraries are a major customer of database vendors, another stakeholder group with a key preservation role. It is here that we can clearly see the interplay between the economics of owning content, the cost of preserving and making it available, and the value equation for the consumers of the content.

NEWSPAPER ARCHIVISTS

Almost all newspapers keep their own archives in some form – even if it consists of stacks of back issues. These archives meet a particular end-user need, which is different from the historians' and

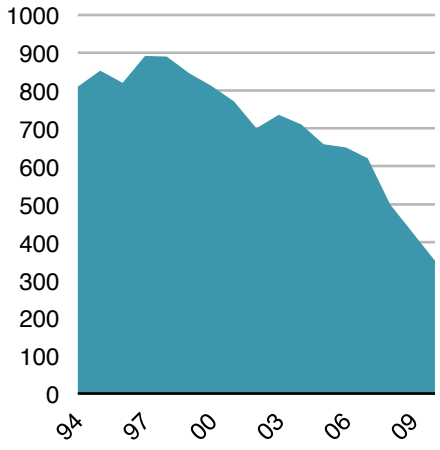
genealogists'. News backfiles exist largely to answer a common newsroom questions: What do we know about this subject? What have we published before, and can it be reincorporated into new coverage? In other words, the questions are about not duplicating effort.

Newspapers that do maintain a formal archives employ a combination of human beings and algorithms, with a growing share going to the latter. Increasingly automated content creation and production workflows have made the capture and, to some extent, the curation of news archives a machine process. However, newspapers are also the owners of some complicated and comparatively old digital assets, as the chart on page 7 suggests. The deep knowledge of their contents, formats, evolution and structure lies with the people who have been responsible for their creation and maintenance, usually the librarian-archivists in the newspaper library.

Discussions of "preservation" with this group usually centers on discussion of page PDFs and capturing websites. But a daily paper of any size is likely to have in its archives one or more of the following digital assets:

- A unique text database of articles and bibliographic metadata and indexing
- A one-off database corresponding to stores of film negatives
- Photographs, web audio and -video stored on aging CD-ROMs
- One-of-a-kind research databases produced for specific journalistic enterprises
- Subject and author indexes to decades of news content

Figure 2. SL News Division Membership, 1994-present



Unfortunately, the numbers of those with intimate knowledge of these assets have been greatly reduced in the deep industry recession that began a decade ago. With the advent of integrated production systems and archives, news archivists have often been the first group to be eliminated through downsizing. (Figure 2 tracks News Division membership in the of the Special Libraries Association. The division was founded in the 1920s for librarians in newspapers and later magazines, broadcast organizations, and nontraditional news settings.)

Whether the algorithms are capable of properly curating

the born-digital news archives for which they're responsible will be an ongoing experiment – and pity the poor library to which the system will one day be handed off.

PUBLISHERS

Professional library and archives organizations from the International Federation of Library Associations (IFLA) to the Society of American Archivists (SAA) have undertaken at least preliminary discussions of born-digital newspaper collections. Notably absent at these confabs are the publishers and trade groups like the Newspaper Association of America (NAA), the international group WAN-IFRA: World Association of Newspapers and News Publishers, and standards group International Press Telecommunications Council (IPTC). While they establish and to some extent regulate data and production standards, they are more likely to engage in the parallel universe of digital asset management (DAM).

"Preservation" in this milieu is usually understood in terms of storage and fail-proof systems. A publisher's assurance by those in charge of the data that the collections are technically sound is going to be sufficient to end the preservation question then and there. If a "preservation" conference gets on a publisher's radar screen at all, he or she is likely to send a representative from the IT department or (if he has one) a news librarian.

The absence to date of publishers in this discussion has been an important handicap to advancing the discussion of rescuing digital news archives. Appealing to press associations will be important in reaching out to this group.

VENDORS

Researchers typically access newspaper backfiles via a third party, either a public scanning project or one of the commercial vendors that produce microfilm for their newspaper clients and/or license newspaper content for resale through various databases (e.g., Readex, NewsBank, ProQuest).

Digital issues are crucial to this group. While they're not formal preservation repositories by any means, they are nevertheless curators of digital assets from newspapers, and maintaining perpetual access to content they've decided has value is their business. As newspapers have downsized their archives staffs, the vendors' role in providing important metadata has increased. They are the de facto archives for many publications and are on the direct receiving end of this highly automated archiving going on by their newspaper clients. Among the vendors, we can see how the complex economics of sustaining digital collections is playing out, as well as the myriad challenges of system-to-system interoperability.

Stakeholders and the Summit

The Newspaper Archive Summit organizers can take credit for putting on a conference that was unprecedented (at least in the author's experience) in its ability to attract participants across all these domains. Many library participants were delighted with the opportunity to interact with people from a variety of information professions – producers, preservers, curators and users in the same room.

Cross-domain teamwork is essential to both the theory and practice of digital preservation, and one of the greatest challenges. Publishers, librarians, vendors and end users will all have to cooperate in planning, creation and management of an infrastructure that will make newspaper content available in the future.

Each group has its own point of departure, naturally, and getting them into the same room to talk about preservation has been an ongoing challenge. But the Summit had some real success with its foray into cross-domain participation. Just getting the conversation going is important.

3. The digital challenges

'I learned a great deal from the actual journalists at the conference about how they see the challenges they face regarding the creation and maintenance of their archival backfiles.'

Issues in news preservation

Before looking at the Summit itself, it is helpful to summarize the problems besetting newspapers in the Digital Age. These are typical of many industries. The possible exception is No. 4, below, which in newspapers may be one of the most compelling examples today of a digital threat playing out in reality.

Setting aside the varying definitions of preservation among participants, there are some fundamental problems that cross all of Summit domains. We'll see, however, that more than any other stakeholder group they impact the publishers.

1. *Relentlessly shifting technologies:* News delivery models that both hasten obsolescence and make any potential solutions an endless game of catch-up.
2. *Holes in the record:* Comparisons of print pages, page PDFs and web pages reveal a tremendous disparity in content from form to form. In the new era, what is the "edition of record"?
3. *The economics of sustainability:* How will preservation be funded when revenue models are uncertain? This doesn't affect just for-profit news organizations but cultural heritage institutions as well, as reduced funding throws into question the sustainability of their digital collections.
4. *What to do about orphan works in digital newspaper archives:* Photos and articles whose copyright status is unclear, subject to potential litigation, and which will never be revenue-producing. Entire newspaper sections have been withheld from digitization because of their copyright problems.
5. *What happens when a newspaper ceases print publication?* Or goes out of business: Are the experiences of the *Rocky Mountain News* and Denver Public Library a roadmap for what is ahead?
6. *The lack of standards and best practices in the industry:* One-off databases and outputs, along with relentless deadline pressures, make it difficult to undertake cooperative ventures in preservation.

7. *The sheer expense:* Digital preservation will require research and development and business-process reengineering at a time when many newspapers are struggling for survival. How they manage it will have a downstream impact on vendors and cultural heritage partners.

URGENCY

And then there is the timing. There are backfiles in strange formats on obsolete media at every newspaper that are either lost or on their way there. There may be web pages "somewhere" on a newspaper's server, but once the newspaper stops linking to them from active pages there is no way to access them unless you have preserved the URLs. They're not necessarily captured by algorithms or archivists.

Beyond this aging-content problem, which is basically one of benign neglect giving way reluctantly to more intentional preservation, there is the microfilm problem.

MICROFILM: GOING, GOING…

American newspapers have deposited microfilm with the Library of Congress for decades as part of the copyright registration process. There is a growing call from publishers to stop sending film and start sending page PDFs, a common by-product of press imaging systems. The expense of having film manufactured is an attractive one to eliminate when the "identical" material is piling up on servers or CD-ROMs. (The fact that CD lifespans are spoken of, misleadingly, in centuries does little to improve microfilm's hopelessly old-school image.)

Such an apparently simple problem is, not surprisingly, intensely complex. But it is certainly an urgent one. So we might add to the digital challenges above:

- *The need to find an alternative to microfilm:* PDFs are abundant and a comparatively stable format, but uniform they're not as produced by various newspapers' systems and local production practice. Finding a standard output will require getting publishers and their standards organizations on board, and this may be an opportunity to leverage some of the knowledge acquired in public scanning projects.
- *The need to keep producing it anyway:* The fact remains, at least for the time being, that microfilm is the most reliable form of preservation. So even if PDF deposit becomes feasible, we aren't ready to abandon film.

These are some of the challenges facing preservers of news content. There will be others. All things considered, it is a very steep agenda for a day and a half of meetings. But the discussion has begun and progress was made, as we'll see in the next section.

4. The Summit: agenda

> ***'My most useful learning experience was that many people face the same questions and challenges and that many voices at the table will make for better solutions.'***

Agenda

Against this complex backdrop of digital preservation in general and newspapers in particular, Summit planners in the fall of 2010 started casting a wide net for participants, panelists and topics. The resulting agenda was ambitious for a day and a half of meetings, but it provided the necessary groundwork for exploring where interests and issues congregated. The final agenda touched on a host of substantive issues, and actual discussions opened up the prospects for further investigation in several key areas that will make for a productive "next time." (The formal agenda is reproduced in Appendix A.)

What were the questions presented by the Summit? A look at the promotional literature spells it out this way:

> *If a newspaper goes out of business, what happens to its archives? If a newspaper becomes available only online, what happens to yesterday's (or last year's) news? Why should we care?*

Participants could expect

> *to discuss the state of newspaper archives and the feasibility and logistics of creating and managing light and dark archives of orphaned and born-digital newspapers.*

Themes

Given the digital challenges and this specific mission, how was the Summit structured?

STAKEHOLDERS

The meetings centered on a daylong series of panel discussions under the general banner, "Who cares about newspaper archives?" Session 1 answered the question from the perspective of historians and genealogists, two groups that derive great benefit from digitized news collections. Session 2 treated the same question from the perspective of local communities, and, again, the takeaway is that a sense of their own history is beneficial to contemporary communities, and the best place to develop it is likely to be the local newspaper archives.

As the first day continued, it was obvious that few people do not care about newspaper archives. However, it is more difficult to dissect how the remaining panels approached the specific question. The overarching question might have been, "We've established that preserving news archives is popular and important, at least in concept. What are the challenges?"

In sessions lasting about an hour each, Panels 4 through 6 addressed some of the complex issues confronting preservers: Issues with digital preservation, some of the lessons being learned in various digitization projects (public and private, U.S. and overseas), as well as general copyright law and publishers' concerns with revenue, or *monetization*.

FUNDING, FUNDING, FUNDING

Monetization was the buzzword of the second day of the conference, and warrants some explanation here.

Preservation, whether analog or digital, is a costly enterprise. Among other things, the motivation to preserve is based on the value of the content, but *value* is in the eye of the beholder. As cultural heritage institutions, libraries and archives have a much more up-front mandate to preserve backfiles than do profit-making entities, whose value criteria are going to be based much more on economics through efficiencies, potential resale and internal research.

In the case of a newspaper, for example, once a publisher decides a digital asset is no longer generating enough revenue to cover the cost of its upkeep, the asset is at risk. In the dire straits many newspapers find themselves in today, even valuable material may be unsustainable, especially if it is not textual or can't be automated – video and blogs, for example. The ability to monetize assets thus motivates publishers to preserve them. (One way to monetize paper assets is to digitize them, although revenue doesn't automatically follow.)

BRAINSTORMING

On Tuesday morning, Summit participants reconvened to tackle the problem of monetization and come up with some actionable ideas for funding preservation of threatened collections, either through jump-starting a few revenue-generating projects or lobbying publishers or foundations to start putting money toward preserving news.

Participants started the exercise by developing an astonishingly rich and diverse list of the potential content hidden in historic newspaper archives. Divided into seven teams, participants continued by developing ideas for how to monetize any of it, or pursue other ideas for bringing in money to support their news collections. (Brainstorming results are found in Appendix B.)

While the feasibility of some of the ultimate ideas turned out to be debatable, the idea was for the groups to form a social network (via LinkedIn) to continue the discussion, if not actually launch a revenue project. Results since have been mixed, the subject of the next section.

5. Summit outcomes

> ***'In some respects, I felt that this conference was stuck in a time warp. The discussions about preservation and access echoed discussions that took place within the library profession in the late 1980s and early 1990s.'***

Searching for commonality

The quote above neatly summarizes the view of a seasoned academic librarian, for whom digitization of newspapers is an old-school topic. Others felt that libraries weren't sufficiently represented and that too much emphasis was on publishers' points of view. Many appreciated hearing other ideas, though:

> *It was enlightening to meet such a broad range of people. Usually I only see librarians at the conferences I attend.*

WHAT IS 'PRESERVATION'

Perhaps one of the most striking outcomes of the News Archive Summit was the wide disparity in understanding of what "preservation" meant at the conference.

Cultural heritage community

Historians, genealogists, librarians and archivists tend to see digitization as preservation, and whatever it means, they want it urgently. The promise and potential of the pre-1923 digitization projects haven't extended to every historical society with old newsprint on its hands, and quite reasonably that is what this Summit was about to them.

Publishing industry

Publishers, on the other hand, have been very careful about digitization for a couple of reasons. One, they as a group were slammed in 2001 with the *Tasini* decision[*], which severely restricted what publishers could do with repurposed electronic content without further compensating the original creators. Freelance contracts have been rewritten to accommodate secondary usage but publishers are still sensitive to this potential liability in trying to monetize backfiles.

Second, there is the dark specter of the digitization of music in the 1990s, where unfettered reproduction of intellectual property basically wrecked an entire business model in short order. The very notion of surrendering digital files to a third party without a guaranteed revenue stream is practically a

[*] *NEW YORK TIMES CO. V. TASINI* (00-201) 533 U.S. 483 (2001) 206 F.3d 161, affirmed.

nonstarter, as was abundantly evident during the Summit – even if that third party is dedicated to preserving the content.

Digital preservationists

Finally, the few people in the room who actually work in the field observed a lack of consensus on a definition of what that digital preservation means. It wasn't so much confusion as it was the lack of a critical mass of ideas to advance the discussion.

Nevertheless, it was remarkable that so many ideas surfaced in this milieu. One statement overheard was

> *What was interesting for me was [how a session] exposed the various stakeholders viewpoints and biases.*

It gives us an opportunity to shape discussion next time, and work on achieving that critical mass.

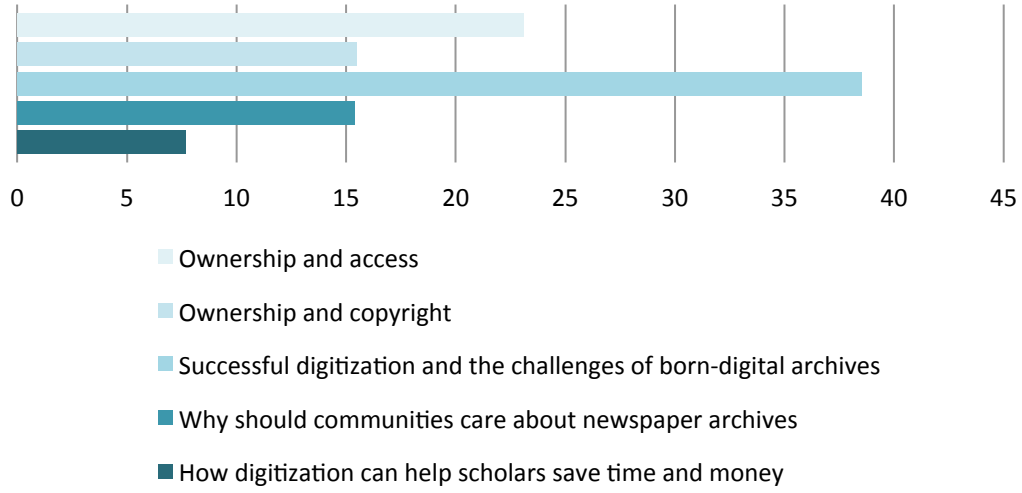
Are the right people in the room?

The responses to this question reveal these biases firsthand. Next time, people want more librarians, fewer librarians, more technologists, more vendors, fewer vendors, more small publishers, bigger players who will get the industry's attention, and so on. In other words, the survey respondents have some special biases related to their local situations and are looking for reinforcements. It will be important to unify stakeholders around some common ideas.

REACTIONS TO PANELS

Participants were surveyed after the Summit on their reaction to the panels. A plurality of interest seems to have landed on the session about digital preservation and practical digitization projects, as Figure 3. shows.

Figure 3. Usefulness of presentations (%)



The next most popular panel was the one on ownership and access, the panel where "monetization" sparked a lively dialogue among participants. The early panels that aimed at understanding the value of newspaper archives to scholars and local communities garnered somewhat less enthusiasm, perhaps because there was so much immediate consensus on the importance of newspapers as a historic resource.

SESSIONS

The dinner on Monday night was a relaxed mixture of idea-generator and icebreaker, with small groups of Summit participants forming and re-forming over different stages of a progressive dinner. Each course offered a provocative discussion topic, such as "You are estate planning for a newspaper. Who would you identify as the guardian?" and "In 50 years where will you find out about your community history if there is no library?" The responses elicited opinions across the spectrum of stakeholders, and the subject of born-digital preservation came up frequently.

Tuesday morning was intended to build on these discussions, but seemed to some participants to be too disconnected from the direct preservation topics of the previous day. It was predicated on the understanding that digitization of newspaper backfiles is a desirable form of "preservation," and the way to motivate additional funding is to establish the underlying marketability of the content. All well, good, and even true – but it didn't directly address the "rescue" of threatened collections or treat it in a practical way.

Although revenue-generation (i.e., economics) is a key component of the preservation equation, whether this was made sufficiently clear, or was too premature in a Summit about "rescue," is reflected

in comments from the initial survey (Figs. 4 and 5). There were only two substantive comments, but one indirectly summed up this problem with differing definitions of preservation:

> *How to generate money for digitization" [the Tuesday morning session] didn't really get at many of the interesting topics raised on Monday - we didn't get into the born digital nearly as much as I'd hoped.*

Figure 2. Effectiveness of Monday night session

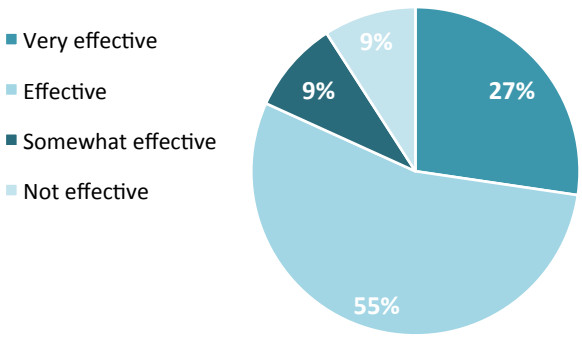
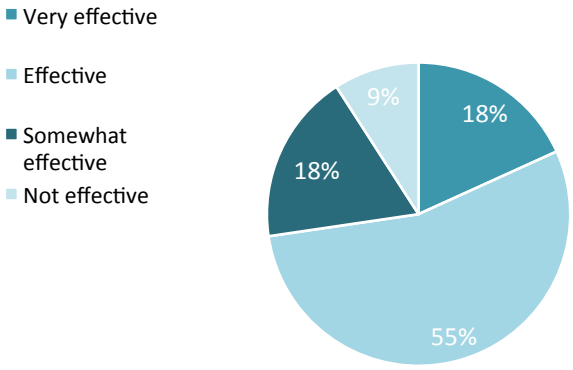


Figure 5. Effectiveness of Tuesday morning session



Another comment touched on the commonality problem, perhaps from a stakeholder in need of more digitization resources:

> *The brainstorming session on Tuesday was disappointing. The conversation assumed that newspapers were already digitized and the ideas generated little on how libraries, publishers, and vendors could work together.*

For an inaugural event, perhaps the interests/biases were too diverse to result in anything immediately actionable. While the brainstorming session was interesting, stimulating and fun, its overall value to the "rescue" theme and concept of a news repository is limited at this point. It makes it all the more important to focus on developing a critical mass of definitions and ideas about preservation. (Brainstorm group ideas are included in Appendix B).

6. Post-summit results

The RJI Survey launched June 24 asked participants what actions they'd taken or changes they'd initiated since the Summit concluded. Many comments were along the lines of "too soon." Respondents were very appreciative of the opportunity to hear from other constituencies; librarians, for example, were interested in what publishers had to say about the sorts of economic issues that seldom confront heritage institutions. Others appreciated hearing about certain digitization projects and planned to incorporate these historic materials in their work.

Another repeated remark among survey respondents touched on their interest in learning more about handling born-digital news. This comment is representative:

> *Hearing the challenges and views of active newspapers and commercial vendors was helpful, but the discussion of archiving digital editions was probably the most helpful and we would like to learn more about how other papers and institutions are approaching the archiving of born digital papers...*

On the other hand, some participants clearly looked at the Summit as an opportunity to gain more information about digitizing their paper holdings, as in this comment:

> *I don't think any important groups were missing entirely, but I think more representatives of large public libraries and state historical societies would be appropriate, as these institutions have both significant holdings of newspapers and a vested interest in their digitization...*

It will be up to Summit II planners to address the needs of these two disparate domains. Will revisiting established practice for digitizing pre-1923 newspapers advance the "rescue" aspect of contemporary and born-digital collections?

Additional comments regarding the outcome:

> *too soon, none yet,*
>
> *long-term effort, still pending ...*
>
> *More interest in a particular format*
>
> *New teaching materials*

And one has to feel some sympathy for this bewildered participant:

I'm not sure results or effects have resulted, except possibly I feel more pessimistic about reaching an understanding than I did before.

Coverage of the Summit

FORBES

Business writer Emily Lambert of *Forbes* wrote about the Summit in the "Trading Places" blog April 18, 2011 (with 1,186 page views). "Finding Money in Newspaper Archives" covered the monetization issue and the problem of sustaining archives with uncertain commercial value.

http://onforb.es/ninxsL

LIBRARY OF CONGRESS

Summit participant Abigail Grotke reviewed the for the Library of Congress' Digital Preservation website on April 21, 2011, "Content Rescue: The Newspaper Archive Summit."

http://1.usa.gov/fnIp7A

MISSOURIAN

Missouri School of Journalism faculty member and executive editor Tom Warhover wrote an article for the *Missourian* explaining the digital preservation problem for ordinary consumers, "Dear Reader: Digital archives don't last: A tale of corruption and crashes." The April 15 article mentions the permanent loss of the *Missourian's* digital archives in 2002.

http://bit.ly/dNZ7pg

7. Looking ahead: Summit II

'Preservation, preservation, preservation.'

Rescue again: What does it mean?

Let us revisit the theme of Summit I and try to parse out a cohesive direction for the next one:

> *If a newspaper goes out of business, what happens to its archives? If a newspaper becomes available only online, what happens to yesterday's (or last year's) news? Why should we care?*

CREATING FOCUS

Focusing on "preservation, preservation, preservation," as one survey respondent suggested, is largely a focus on educating participants in what the word actually means. This is no small matter. As Summit I made clear, different definitions and mindsets about preservation contributed to a certain amount of confusion and distraction from the theme as participants absorbed the proceedings through their own biases and understanding.

A thorough investigation of competing definitions can only help keepers of digital collections do a better job, whether they're publishers or historical societies. Whether digitization is an acceptable form of preservation is not entirely settled in the library community, and exploring the nuances will benefit everyone, especially as the problems of multimedia formats become better known.

And in the domain of digitization, is there a clear set of priorities?

- Is there material at such risk that digitization could indeed constitute a rescue?
- Are there digital formats at greater risk than others?
- Should rescue of digital collections take priority over orphaned newsprint?
- Are there ways among the Summit stakeholders to determine priorities?
- In the words of one participant, "What is needed most now, and what can wait?"

To those ends, here are a couple of urgent topics:

ARCHIVES OF DEFUNCT NEWSPAPERS

The national and state scanning projects under the National Digital Newspaper Program umbrella (http://www.loc.gov/ndnp/) have been working in this field up to the 1923 copyright wall. This has resulted in a deep body of knowledge and best practices for digitizing physical copies and applying metadata. It serves the preservation community very well up to about 1985, when news archives

become mostly digital. At that point, the interests of publishers and the cultural heritage community in preservation start to intersect.

That suggests a new theme,

What have we learned from digitization that will help us with defunct newspapers?

Clearly that isn't a catchy phrase, but it's important to find one that will help achieve focus and critical mass.

ONLINE ONLY

What does happen to yesterday's (or last year's) news in a born-digital environment? Interrogating that question is the essence of digital preservation itself and its myriad considerations, as listed on page 10:

- Technological change
- Identifying the authoritative "edition"
- Economics of sustainability
- Orphan works and copyright
- What to do with a defunct paper that exists only electronically
- Need for industry standards
- Continued role of microfilm in a born-digital world

Individually these are all topics of discussion for Summit I, which we'll see in the concluding section.

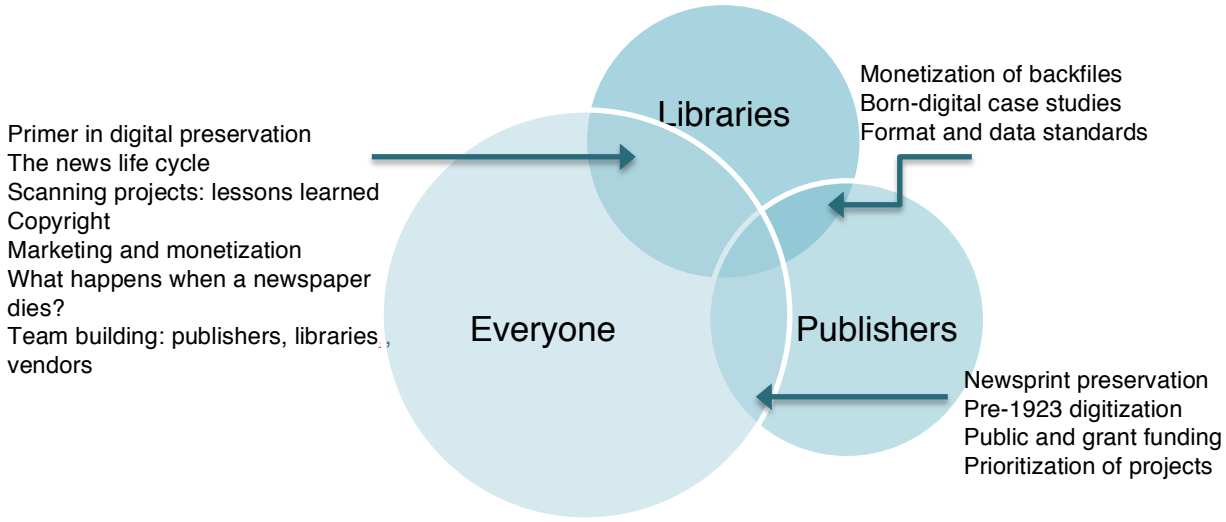
WHY SHOULD WE CARE?

There is not a stakeholder group from Summit I that wouldn't care about these topics. The answer to the question, "Who cares?" lies in the fundamental problems of digital preservation: How publishers and heritage institutions alike will be able to sustain their digital collections through times of continual change and uncertainty. Convincing publishers in particular is a challenge, but the Missouri School of Journalism and Reynolds Journalism Institute are better positioned to do this than the cultural heritage/preservation/library community by itself.

Seeking common interests

A brief analysis of some of the topics that came up in Summit I suggests that there are more areas of shared interests (Figure 6) than otherwise:

Figure 6. Areas of commonality



It might be worth exploring a two-track approach for Summit II with sessions of interest to libraries and archives and others for publishers. For example, a genealogist or historian may not be interested in a discussion of data standards with news industry experts and Library of Congress representatives, but all might be interested in discussions about the news life cycle. Two tracks would be a way to keep the panel discussions focused and not overly general. The trick is encouraging cross-domain understanding and participation.

Topics

REACTIONS FROM SUMMIT I

The RJI survey (Appendix C) asked participants to suggest topics of discussion for Summit II. As with other questions, the answers reflected the participants' particular background and interests. Some wanted more presentations from publishers, others from librarians. One common area of agreement, however, was interest in hearing more about preservation. Whether that meant more about digitization projects or specific preservation practices and case studies was less important than the desire simply to know more.

While some felt that the topic of monetization was not particularly helpful (perhaps a reflection of its sometimes contentious tone), others thought additional discussions between publishers and the academic market (where actual monetization has been going on for a long time) would be fruitful for all sides. There was also an expressed desire for more discussion of potential partnerships between producers and libraries, libraries and vendors, producers and vendors. Fostering these can only enhance the capability of creating a sustainable effort, whether an annual Summit or an actual repository.

TOPICS FOR SUMMIT II

While these suggestions are excellent as far as they go, there is also a great deal of research under way in digital preservation worldwide, some of which can be potentially useful to efforts to preserve digital news. The following topic suggestions combine user responses with some of these high-level themes. These, in turn, may suggest additional participants and panelists.

- *Preservation 101.* Reaching a common understanding of what preservation means in various contexts, and especially the risks to any digital collection.
- *Understanding news archives:* Why do publishers keep archives in the first place? What purpose do they serve, and how does it jibe – or conflict – with the public benefit of permanent, historic collections? What is a journalist's view of the archives?
- *Identifying at-risk collections:* Once the components of digital risk are understood, how do they apply to what publishers keep? Are text systems robust while photojournalism is under threat? Web pages, audio, video and the other mix of news channels – tweets, blogs and social media? Can we make evaluations now on behalf of future users? Can we learn to prioritize?
- *Economics:* This might be called (tongue in cheek) "Beyond Monetization." A hard look at the burdens, benefits and free riders in this equation. The *sustainability* of digital collections is a major topic around the world and there is starting to be a lot of literature on the subject.
- *Copyright, embargo and the concept of 'dark archives':* Economics is intimately intertwined with copyright. Are there workarounds to publishers' concerns about intellectual property? What are the allowable preservation actions under copyright law?
- *Trustworthy repositories:* Much work has been done in the last several years (by the Center for Research Libraries in particular) on what constitutes a reliable preservation repository. What does "trustworthy" mean? How many digitization projects out there know what to do to keep their scanned collections viable? Is there a model for a news repository? If Missouri undertakes a dark archive, what will it look like?
- *What about page PDFs:* What do all the millions of scanned pages and extracted articles have to tell us about born-digital preservation? PDF frontiers, what have the scanning projects taught us about PDF deposit?
- *Carrots, sticks, and microfilm:* If publishers want PDF deposit instead of microfilm, how do we advance the discussion? What can two decades of the National Digital Newspaper Program tell us? There are a few PDF deposit projects ongoing in various parts of the world. What are the useful models out there? What will a U.S. testbed look like?
- *Standards and Editions of Record:* PDF deposit will require newspaper production standards that may exist on paper but are remote in practice. What are they? Does 150 years of wire service experience have anything to teach? How do we understand what to capture?

- *Roles and responsibilities:* Digital preservation requires actions by a spectrum of "agents" from creators today to end-users 50 years from now. How can stakeholders be recruited and represented? How can partnerships be fostered? Where should we start? Systems vendors? Press associations? Journalism schools? Trade groups?
- *Anatomy of a rescue:* Instances are growing of papers going out of business, or less radically, opting for electronic-only publication. What happens to the archives? Are there enough case studies for a critical mass? Can we start identifying what is truly at risk?

One participant made this cogent remark:

> *Papers are dying - we should talk about trying to save their archives before it is too late.*

The extent to which any of these topics touches on the urgency of this statement will drive the mission statement of Summit II.

The role of recruitment and RJI

One participant remarked that it was nice to be able to interact with non-librarians. This touches on a curious situation with preservation. Innumerable conferences, convocations and committees in the library profession have grappled with digital preservation, and the problem is largely seen, and often *dismissed*, by the rest of the world else as a "library problem."

Publishers are more likely to see it as a "computer problem." Few newspaper CIOs are familiar with what we might call the "cultural heritage perspective" of digital archives, and understand it instead as a storage and backup problem to be solved. When they reassure their publishers that their systems are robust and failsafe, publishers tend to leave it at that – unaware, for instance, that they may be losing a fraction of their digital archives every year to data corruption. (Even publishers who grasp the situation are likely to set it aside in the face of bigger problems.)

Publishers are critical to the success of a newspaper preservation effort and are the hardest to enlist. Summit I broke new ground by attracting attendees from outside the usual domains of the news library and IT department.

The presence of professional journalists/professors at the Reynolds Journalism Institute is likely ensuring more success than even an invitation from the Library of Congress. But it will take more concerted involvement by Missouri and other journalism school faculty than was the case with Summit I. Enlistment has to start from within. Social media makes this a relatively painless effort, at least to start.

The promise of more discussion of rescue and preservation, and the promised presence of publishers, will in turn encourage more enthusiasm from the cultural heritage community. In the next section, we'll look at how social networking can encourage participation in Summit II.

Conclusion: Building on Summit I

Next time?

There are clearly many opportunities for bettering the understanding of stakeholders and defining their core values on the way to developing a critical mass of interest and concern in newspaper preservation. This presupposes that a Summit II is an important and worthwhile undertaking and has proposed some approaches for "next time."

Meanwhile, there are opportunities to build on the unique gathering that was Summit I with its diversity of communities. The biggest challenge will continue to be to keep participants engaged, at least peripherally, between now and when the group reconvenes. Ideally, they'll have a stake in putting Summit II together. Steps might include:

SOCIAL NETWORKING

There was some debate at the conclusion of the Tuesday brainstorming about staying in touch, centered on whether a social network site would accomplish that goal, and whether Facebook or LinkedIn would be the right platform. The majority seemed to side with LinkedIn as the more business-oriented of the two, and a general group was created, along with individual color-coded groups for each of the brainstorming teams. The idea was for them to continue the work begun with the brainstorming and to try to bring some of the ideas to reality.

The color teams in the meantime haven't generated much activity, and some of the comments from the RJI Survey suggest that it would be premature to expect any. Given the larger issues set forth in this document, perhaps they're best set aside at this point. Focusing on the general group as a communications channel can resume with dissemination of this white paper for follow up, and as interest is kindled in Summit II, additional information – and requests for input and help – can be pushed to members.

As we've seen, Summit I participants should be encouraged to recruit colleagues through their own contacts. After Summit I, participants had some ideas about whom and from what domain they'd like to see participating next time. Through social networking, this ought to be more attainable than it would have been even a few years ago. The following table are some initial suggestions, a few with names

attached, of people who might be recruited. The online network of "I know someone who knows someone who knows x" could result in some terrific participants.

The following table is a kernel of suggestions, and far from comprehensive but a starting point.

Topic	Suggestion
Building light and dark archives in the Digital Age	Martin Halbert, MetaArchive
Economics of preservation	Brian Lavoie, co-chair of Blue Ribbon Task force
Newspaper production systems, especially where PDF deposit has been under way (e.g., Singapore)	CCI Denmark, dominant company
Raju Buddharaju of Singapore Library Board	Project of electronic newspaper deposit
Standards	Representatives from NAA and IPTC (Michael Steidl from IPTC)
News librarians and leadership	Eli Edwards, chair of news division, SLA; David Cappoli, recent SLA president candidate and former division chair; Justin Scroggs, former chair, librarian, licensing officer at Newsbank
Journalism professors/news librarians	Nora Paul (U Minn.)
	Kathleen Hansen (U Minn.)
Journalists with perspective on preservation	Michael Miner, columnist for the Chicago Reader
Library of Congress and NDNP	Martha Anderson, Mark Sweeney
Leaders, vendors or practitioners from Digital Asset Management	Numerous groups in LinkedIn occasionally take up preservation-related issues. Invitations could be issued through them
IFLA Newspaper Section	Frederick Zarndt is current chair; section will be meeting in August 2012 in Finland

Educating stakeholders

Because many of the Summit I participants, especially from journalism, may be unfamiliar with preservation issues, a carefully chosen selection of current research literature could be pushed via Linked in or Summit II website.

Identifying relevant articles would be an ideal project for a journalism graduate student, or a team of students, including library science or even computer engineering. (This need not be confined to students at Missouri, especially if Summit II is held at another university.) The goal would be to provide a basis of familiarity so that the various definitions of "preservation" would be clearly understood across the stakeholder spectrum.

This would serve the additional goal of providing regular updates and keeping the LinkedIn social network active. If any of the participants are regular bloggers, posts in this area could similarly be pushed to the group and comments invited.

Advancing the Research

Provocative questions could be raised in the group forum online. One thing that the group can do as a whole would be to start trying to identify actual incidents of loss.

Without much investment of time or money, group members could start soliciting questions from press association members about collections that they know to have been damaged or disappeared, whether analog or digital. This is actually a comment that frequently comes up in digital preservation at large: that the potential loss is indeed great at the theoretical level, but hard instances (known as "actualizations") are few and far between. A website or wiki with contributions and updates could become compelling at a national level, and interesting to do.

A graduate student or team from the Journalism School might also revisit the survey of news archives undertaken in 2005-06 for the international archives preservation consortium InterPARES (International Research on Permanent Authentic Records in Electronic Systems). Using that set of questions and results as a benchmark, Summit II could attempt to quantify the at-risk collections, or go for a qualitative analysis of risk in light of the current dismal newspaper economics. What does it mean to news collections to be preserved via algorithms?

The end – for now

These are just a few suggestions to look at when planning for a Summit II. This document is not intended to be an end-all of the discussion but rather to stimulate new directions. Unlike that poor respondent who felt more pessimistic about preservation after April gathering than he did before, we want to go forward with a lot of optimism and resolve.

Appendix A: Agenda

Summit information, including videos, is archived at:

- http://www.rjionline.org/events/newspaper-archive-summit
- https://mospace.umsystem.edu/xmlui/handle/10355/10617

Date	Session
Sunday, April 10 Dinner speaker	Aaron Presnall, Director of Studies, Jefferson Institute
Monday, April 11 Keynote speaker	Martha Anderson, Director of Program Management for the National Digital Information Infrastructure and Preservation Program, Library of Congress
Saving Time and Money: How Newspapers Digitization Can Help Scholars	Moderator: Earnest Perry, Associate Professor, Chair, Journalism Studies, Missouri School of Journalism. Panelists: Bob Allen, Professor, Drexel University; Bernard Reilly, President, Center for Research Libraries; Chris Cowan, Vice President, Publishing, ProQuest ; Patrick Cox, Associate Director, Briscoe Center for American History, The University of Texas at Austin
Why Should Communities Care about Newspaper Archives?	Moderator: Debra Mason, Professor, Journalism Studies and Director, Center on Religion & the Professions. Panelists: Remmel Nunn, Vice President for New Product Development, Readex division of NewsBank; Joe Hight, Director of Information and Development, *The Oklahoman*; LaDonna Garner, Historic Preservationist and Certified Genealogist; David E. Rencher, Chief Genealogical Officer for FamilySearch
Lunch speaker	Nick Townend, Head of Digital Operations, The British Library
Digitization: Successful Projects and the Challenge of Born-Digital Newspaper Archives	Moderator: Vicky McCargar, Journalist, Archivist, Digital Curation Consultant. Panelists: Martin Halbert, Dean of Libraries, University of North Texas, co-founder of MetaArchive; Frederick Zarndt, Chair IFLA Newspaper Section, Consultant for Digital Divide Data; Abbie Grotke, Web Archive team Lead, Office of Strategic Initiatives, Library of Congress; Leigh Montgomery, Librarian, *Christian Science Monitor*; Jim Draper, Vice

	President and Publisher, Gale/Cengage
Ownership and Copyright Issues	Moderator: Richard Reuben, James Lewis Parks Professor of Law. Panelist: Caroline Pinkston, JD Candidate '11, MU Law School
Ownership and Access Issues	Moderator: Brian Steffens, Director of Communication, Reynolds Journalism Institute. Panelists: Christopher Gill, Founder, Heritage Microfilms and NewspaperArchive.com; Randy Weissman, News Administration Editor, *Chicago Tribune*, Mizell Stewart, Editor, Scripps and *Courier & Press*
Evening	Facilitated brainstorming over dinner
Tuesday, April 12t	
Developing an Action Plan for Preservation and Access	Moderator: Stephanie Padgett, Assistant professor of strategic communication.

Appendix B: Brainstorm groups

Developing an Action Plan for Preservation and Access. These lists and notes were collected during the brainstorming sessions April 12 and sent to Summit participants.

What is in a newspaper archive?

Recipes & menus

Obituaries

Birth announcements

Weddings & anniversaries

Marriage/divorce notices

Public Notices

Local sports scores

Photos [see below]

Event calendars & dates

Schedules

School lunches

Church events

Travel guides

Eyewitness accounts

Voting records

Election results

Games

whist; acrostics; word scrambles; math problems; bridge; chess; jumble

Critiques and reviews

entertainment reviews

car reviews

Ads

Names: alumni

Letters to the editor

Poems; early fiction

Before they were famous

TV, radio, movie and theatre listings

Real estate sales

Estate sales

Probate records

Lottery numbers

Military casualties

Stock market closings

Business history

Women's pages

Society pages

Fashion; art; style

Maps

Weather

Astrology

Pet of the week

Veterinarian columns

Pets for adoption

lost/found

Animal control

Business records of the newspaper

What was in a newspaper archive (100 years ago)?

Railroad schedules

Shipping news

Manifests

Police records

Hotel registers

Who is visiting whom and from where

Serial fiction

Reprints of stories

Veteran reunions

Poetry

“Post boy poems”

Post office: mail waiting

Patient listings; illnesses

Epidemic deaths

Institutionalizations (sanatoria)

Candidate endorsements

Candidate speeches

Etiquette: Miss Manners

Agricultural news

Commodity prices

Slave auctions

Court trial transcripts

Livestock auctions

Ship manifests

Cartoon (political and comics)

Kids’ papers

Horses; racetracks

Patient listings; illnesses

Mosquito abatement (stay indoors)

Cattle clatter

Honor students

Graduations

Crime stoppers

Arrests; mug shots

Citizen-submitted stories (user generated)

1st person travelogues and accounts

Labor union info

What is in a newspaper photo archive?

Disasters

Buildings

Food

Big tomatoes

Bagged deer

Big fish

Trapped animals

Events

Rallies

Sports

Obits (old photos of deceased)

School kids

Weddings; anniversaries

Galas; society pages

Cities

Skylines

Babies

5-generation photos

PR material; celebrity handouts

Weather Photos

Maps

Planning and zoning

Publicity

Inserts

Dairy Princess

County fairs

Pre-photography images: sketches

"Yard of the week"

Garden club

War Photos

Criminals

Historical postcards

Satellite/aerial photos

Scientific illustrations

Household items (e.g., gadgets; new appliances)

In support of the photos that do run:

Outtakes

Negatives (variety of formats; all versions)

Photos from other sources

Doctored photos; censored photos (e.g., painting on an undershirt)

Full photos vs. cropped photos (editing)

IDEA GENERATION

Top [2-3] ideas per table are below. Highlight indicates idea that captured the room's imagination and/or that was further developed into proposal. [Diagrams are omitted.]

RED TEAM

1. Sports Archives
    a. Photos; stats
    b. Domestic or international (how scale)
    c. High school; collegiate; professional
    d. Consumer interest is high
        I. Fantasy sports
        II. Soccer
2. Photopedia
    a. Crowd source annotation of back photos
    b. Invite public to provide context
    c. Historical photos; regular people

BLUE TEAM

1. Corporate/historical perspective to lifestyle issues
    a. Betty Crocker recipe archive (get General Mills funding)
2. University Archives
    a. Reporters notes
    b. Work together to tell the story
    c. Historical look at a journalist's work (behind the scenes)
3. Monetizing political cartoons and their history
4. 3rd party foundation partners (process)

GREEN TEAM

1. Building communities of interest (with payment model)
2. Birthday related
    a. Your date; your location
3. Hobbies
    a. Building, making things (getting back into "making" things; chicken coops; patterns; etc.)
    b. Personal family history

4. Photos, maps, timelines
5. Clip Art files

PURPLE TEAM

1. Business/industrial history
    a. Data modeling
    b. Looking at business through lens of history
    c. Before entrepreneurs were famous
    d. Profiles
    e. Business starts/openings/closings
2. Power of personal names
    a. Vital records and beyond
    b. Letters to the editors
    c. Indexing personal events
    d. Fill in the missing link
3. Education
    a. School districts
    b. Administration moves
    c. Sports teams
    d. What they were spending $$ on
4. Cold Cases
    a. Unsolved mysteries in local areas
    b. Comments; discuss

ORANGE TEAM

1. "Our Book"
    a. Customized books; editors pre-select topics
    b. Themes: industries; neighborhoods; famous cities; giant vegetables; presidents or pols
    c. Features: print on demand; customized pages
2. "My Book"
    a. Personal clipping service
    b. Assemble material into a template
    c. Features: print/web/eBook on demand
    d. Tiered cost

PINK TEAM

1. Foot note-like model
    a. The photos that didn't run
    b. Annotate the photos
2. Civil War History
3. Adopt-a-year/day/birthday
    a. Sponsor digitization
    b. Company anniversaries

YELLOW TEAM

1. NEH challenge grants or tax deduction/credit to local media to digitize backfiles (i.e., stuff in the copyright era)
    a. Incentivizing digitization of backfiles
2. Federal tax deductions and credits for exposing metadata by commercial aggregators
    a. Expose what you have (create efficiencies)
3. University development
    a. Alumni offices
    b. Partner with them to digitize news related to universities; alumni

IDEA DEVELOPMENT

RED TEAM

Team Members: Jim Draper; David Rencher; Ann Riley; Mizell Stewert III; Nick Townend; Jocelyn Wehr; Kristopher Wheeler

Idea: White paper from state and national press associations to connect newspaper publishers with libraries, vendors and other institutions committed to preservation.

Economic Model: Future distribution and commercialization. Affiliate opportunities to build applications on top of the asset: e.g., Photopedia, sports.

Barriers/Challenges: Publishers' and content owners' expectations. Preservation through digitization is first priority.

Key Players: Press associations; journalism schools; vendors; historical societies/libraries.

BLUE TEAM

Team Members: Deborah Cheney; Abbie Grotke; Martin Halbert; Vivian Hay; Gerald Hirsch; Aaron Presnell; Randall Weissman

Idea: National Preserving Our News Heritage Program (NPNHP): A partnership between newspapers, archives & libraries, and funders to preserve the notes of reporters and journalists.

Economic Model: Proposed three way partnerships between newspapers & libraries, with outside, 3rd party funding to build sustainable archive programs with appropriate legal safeguards. Funding could be a tax break for newspapers and journalists; grant funding for libraries and newspapers (IMLS.gov)

Barriers/Challenges: Embargo period necessary. Getting reporters to participate. Incentives are key for newspapers.

Key Players: Newspapers: provide content from archives. Universities/Libraries: apply for grants, house the archives. Funders: sponsor the national program, get it started. "National Sponsor": NEH or LC or someone could promote the NPNHP (NDNP Model)

GREEN TEAM

Team Members: Martha Anderson; Michael Church; Chris Cowan; Patrick Cox; Joe Hight; Ron Larson; Earnest Perry; Erika Van Vranken; Deb Ward

Idea: Building communities of engaged readers/participants around newspaper content that pertains to hobbies, arts, crafts, making and building things. Begin with gardening and cooking.

Integrate weather, localization of gardening practice – assumes large volume of content broadly representing the national. Some combination of the cooking and gardening, "building chicken coops." Questions of how users can contribute to the knowledge base recommendations – what is the date scope of the data? Market through current gardeners' web sites, farmers markets, the "eat local" movement, organic growing. Recommend that it be up to date.

Economic Model: Subscription model. Some free use to create incentive to subscribe. Some ad support (allow for local advertiser as well as national)

Barriers/Challenges: Visibility (how to market and promote); accumulating content providers; distributing revenue to content holders as well as costs of content management; how to capitalize this (foundation; non-profit); need infrastructure to accommodate subscription base; may need some expertise.

Key Players: weather.com. Holders of digital newspapers: further content through press associations & vendors. Agricultural schools.

PURPLE TEAM

Team Members: Kent Ford; Daryl Garwood; Leigh Montgomery; Remmel Nunn; Stan Schwartz

Idea: Business / Entrepreneurship / Industrial / Corporate Information from Newspapers

Economic Model: Mix of support for the backfile; corporate foundations; commercial aggregator; scholarly projects for economic development

Barriers/Challenges: Archiving backfile, getting partners. Contributors. Old copyright issues. Privacy issues. Buy-in by publishers. Ethical challenges (impartiality).

Key Players: State economic offices. E-publishers. Public libraries. Entrepreneurship/business schools. Tech sector. LOC. Newspapers. Aggregators. Photo archive entities. State press associations. Association of State Press Associations.

Going Forward: Material is being gathered digitally on state level, could be scalable. Public notice issue – digital archive of these. Scalable to other states. Searchable – with list of companies.

ORANGE TEAM

Team Members: Bob Allen; John Dougan; Robin Hubbard; Anselm Huelsbergen; Margaret Knecht; Vicky McCargar; Tom Warhover

Idea: "My Book": "personal clipping" service. Revenue from archives at smaller publishers.

Economic Models: Consumers and businesses access digital archives to create personalized digital or print products. Revenue support preservation and expansion archives.

Barriers/Challenges: Vendors and revenue sharing with digitized backfiles – who owns it? Software development. Copyright. Critical mass of participants.

Key Players: Press trade groups and associations and stewardship organizations. NewsBank (Chuck Palsho, President). NDNP – LOC (Mark Sweeney). Small press, public projects (digitization)

PINK TEAM

Team Members: Rachel Brekhus; Gail Gibson-Ranallo; Sue Kellerman; Michael Meiners; Kathleen Schweitzberger; Frederick Zarndt

Idea: Adopt an issue (adopt a birthday, anniversary, founding year [corporation]…). Digitization and p.o.d.

Economic Model:

- Library borrows duplicate master's microfilm from publisher/Proquest
- Library digitizes/has digitized issues on demand (1 reel at a time) and OCR's it.
- Library puts it in a digital archive, adds donor names
- Library and publishers market a tiered sponsorship to public (individuals, corporations, institution)
- Find a way to incorporate annotation, correct-a-newspaper volunteer work
- Charge a little to get your name associated with the record?

Barriers/Challenges: Timing/workflow – like ILL (?). Getting people to contribute. Dividing out material for which someone else owns. Copyright – charge extra (?). Marketing. Publisher resistance. Figuring out how to incorporate annotation (Footnote?). Really old issues may be harder to get sponsors initially.

Key Players: Newspaper itself; cultural heritage institutions holding filmed titles. Associations: state newspaper association, historical association. Vendor, library of state historical society to do scanning. Standards consortia: e.g., library division of SLA, SAA (allows distributed work). NDIPP. Volunteers. Crowd funders (e.g., Kickstarter.com). Frederick Zardnt (IFLA, knows about Footnote)

YELLOW TEAM

Team Members: Ian Buchanan; Christopher Gil; Craig Horn; Nina Johnson; Bernie Reilly [and one more…]

Idea: Federal/foundation support for archiving work and digitization of backfiles by local media organizations.

Economic Model: Tax incentives to cover local media/newspaper archiving costs. Limited term like "cash for clunkers" and must be spent locally, not offshore or at universities.

Barriers/Challenges: Pulling together / consensus building around diverse partners with diverse/competing interests. Copyright /fair review streams. Public money for public good.

Key Players: Fed government. Foundations: Macarthur, Knight, Carnegie. Advocacy orgs: NAA, AAU, Press Associations (state & regional), NNA (small). Media companies: e.g., Tribune. Universities. Research Libraries.

Appendix C: Surveys

Initial follow-up survey - April 18, 2011

1. OVERALL HOW EFFECTIVE WAS THE PROCESS IN GENERATING POTENTIAL NEW IDEAS FOR NEWSPAPER ARCHIVES?

Very effective	Effective	Somewhat effective	Not effective
53.8	30.8	15.4	0

2. WHICH PRESENTATION WAS MOST USEFUL FOR YOU?

1. Saving time and money: How newspaper digitization can help scholars?	7.7
2. Why should communities care about newspaper archives?	15.4
3. Digitization: Successful projects and the challenges of born-digital newspaper archives	38.5
4. Ownership and copyright issues	15.5
5. Ownership and access issues	23.1

3. HOW EFFECTIVE WAS THE MONDAY EVENING EVENT IN SETTING THE TONE FOR TUESDAY'S BRAINSTORMING SESSION?

Very effective	Effective	Somewhat effective	Not effective
27.3	54.4	9.1	9.1

4. HOW EFFECTIVE WAS TUESDAY'S BRAINSTORMING SESSION?

Very effective	Effective	Somewhat effective	Not effective
18.2	54.4	18.2	9.1

5. DID THE PRE-EVENT AND TRAVEL PLANNING MEET YOUR NEEDS?

Yes	No	N/A
76.0	0.0	23.1

Other responses

Being able to attend online was invaluable!

Attended only 2 sessions.

6. DID YOU FEEL THE FACILITIES WERE ACCEPTABLE FOR THIS TYPE OF EVENT?

Yes	No	N/A
100.0	0.0	

7. WERE RJI STAFF ATTENTIVE TO YOUR NEEDS?

Yes	No	N/A
100.0	0.0	

Other responses

very very helpful. kudos to the RJI staff!!

8. WOULD YOU ATTEND ANOTHER EVENT AT RJI?

Yes	No	Maybe – depends on topic
92.3	0.0	7.7

Other responses

The food and facilities were exceptional. It was all very well organized and seemed to run very smoothly. The brainstorming session on Tuesday was disappointing. The conversation assumed that newspapers were already digitized and the ideas generated little on how libraries, publishers, and vendors could work together. The conference was dominated by current publishers/editors and for-profit vendors. How to monetize near-bankrupt newspapers and the Chicago Tribune dominated and prevented discussions on historical, orphaned content.

The true test will be what happens later - if we can work together to solve some of the problems (or not)

I though Monday night was wonderful, but the discussion topics on Monday didn't have any correlation with Tuesday's discussion really. While Tuesday was interesting, the "how to

generate money for digitization" didn't really get at many of the interesting topics raised on Monday - we didn't get into the born digital nearly as much as I'd hoped. 4/19/11 3:22AM

Very impressive with all the speakers

well done and much needed.

The Summit was well hosted.

Very stimulating!

Reynolds Journalism Institute Survey - June 24, 2011

Comments are included as written by participants, including typos

Newspaper Archive Summit Impact Survey Results

What were the most useful learning experiences or takeaways from the workshop/symposium/meeting?

The legacy of print and news, usefulness of newspapers in archiving, genealogy, etc.

Learning about the technologies other companies and institutions use for content management as well as seeing the similarities and differences in the problems they are facing.

In some respects, I felt that this conference was stuck in a time warp. The discussions about preservation and access echoed discussions that took place within the library profession in the late 1980s and early 1990s. It was fascinating to me to see so many "stakeholders" at this conference (newspaper publishers and vendors, for the most part) who didn't have any knowledge of the United States Newspaper Program, for example. So to me the biggest achievement of this symposium was bringing these parties into the discussion.

that many people face the same questions and challenges and that many voices at the table will make for better solutions

I learned a great deal from the actual journalists at the conference about how they see the challenges they face regarding the creation and maintenance of their archival backfiles

I really enjoyed the talks by Robert Darnton [sic] and the Southern History guy (can't remember his name...), but probably the most useful ones involved the Project Nine presentation (the historical collection).

There was more interest in how to milk more money out of material and very little interest in archiving as a research tool and as a publication of record

The personal contacts and network that hopefully developed

I was especially happy to meet a broad range of stakeholders who work on preservation issues for newspapers.

Cross-section of people sharing their views

Cross-sector discussions of the economics of news archiving between publishers, librarians, aggregators. Archiving will all boil down to costs and business models. As a librarian it gave me a better sense of the publisher's viewpoint.

It was enlightening to meet such a broad range of people. Usually I only see librarians at the conferences I attend.

Thinking about the nature of digital news, the importance of news regardless of format, and considering all of the users / stakeholders that need for it to be preserved.

Hearing the challenges and views of active newspapers and commercial vendors was helpful, but the discussion of archiving digital editions was probably the most helpful and we would like to lean more about how other papers and institutions are approaching the archiving of born digital papers. The dinner was very effective in getting people to open up and talk to other participants they would not normally talk to so we really liked that aspect of the conference.

It was very interesting to hear directly from the newspaper industry/publishers/newspapers. I learned we have a lot to hear to get inside their heads. More one-on-one with the newspaper industry. They need to understand the academic market better.

The other media and how they were handling the transition to online. -- The public-private partnerships and the possibility that they can be strengthened in the future. -- It was also interesting to hear the perspectives from the national and British archives, etc. -- An extremely beneficial and well-planned conference overall! I've received several comments based on my panel alone.

Most useful learning experience: It may be possible to cooperate / collaborate with newspaper publishers but only if they are willing to adapt to changing times.

What changes or improvements did you make as a result of what you learned?

Reviewed my own family photos for possible digitizing, became more aware of e-news

Newz Group views its archive of recent content as being more important for researchers, archivists, historians, etc. than we once realized.

No response

None yet

No response

It's just keeping me aware of the importance of getting scholars together to create open access scholarly collections (especially for older materials--out of copyright).

more concentration on the role of newspapers for academic and historical research

Ideas for focusing the next Summit

I had several follow-ups after the meeting.

I was not in favor of using LinkedIn as our communication post-meeting. I don't think it is working.

We are pursuing the federal incentives for news archiving idea that was floated during the last session.

I have been in touch with a few other attendees and we are collaborating on preserving newspapers.

Met a vendor on the first night and we are discussing ONE possible project in ONE format. Advanced some ideas, which generated discussion among our professionals, and was also cited in a Forbes.com magazine article.

Nothing yet.

I'm not sure the Summit resulted in improvements or changes in my own environment, but I certainly learned a lot and have thought about the Summit and what I heard many times since.

We are exploring new ways to handle our archives and possibly digitize more of them in the future. We are also exploring the team's ideas/recommendations for possible usage.

Not applicable

What results or effect have those changes or improvements made?

TBA

We've been looking to tap into the academic/research demand for services like ours. Check out http://newzgroup.com/newzprompt.htm

No response

None yet

No Response

Right now, not much, except that I plan to look for more projects that I can use for teaching, because these projects are a great alternative to materials than textbooks. I enjoy creating, and having students create, projects that 'bring things together,' online, and it's sort of confusing to implement these.

None that have drifted to the bottom of the pond....end users

Too soon

No response

None

Still pending. This is a long-term effort.

The more we work together the better. Duplication of effort is a waste of resources.

More interest than in recent years on preserving a particular format.

Nothing yet

I'm not sure results or effects have resulted, except possibly I feel more pessimistic about reaching an understanding than I did before.

We are already beginning a process to turn archival content into relevant and useful content today. We are also exploring options as to how to monetize our archival content and preserve more of it for the future.

NA

What topics would you like to see addressed in Summit II?

Project Gutenberg and similar programs.

Funding for archiving projects. Newz Group has a lot of capabilities and good relationships with publishers. We'd like to be a part of archiving projects but, of course, we need money to do so.

I would like to see more explicit discussion of what the library profession has been doing all along to provide access and preserve newspaper content. Having someone from the U.S. Newspaper Program, as well as representatives of the National Digital Newspaper Program (which of course is concerned only with pre-1923 content), would be useful. I would especially like to see the former directors of the state newspaper projects under USNP participating. They are the ones who have been having these conversations for the past 25-30 years and who have the best grasp of the problems. It seems to me that we already have a model for addressing many of the issues

raised at the Summit if we can recruit some of the same leaders and convince NEH and other agencies of the need for their support.

preservation preservation preservation

More on how libraries can partner with newspaper and commercial publishers to digitize the backfiles of orphaned newspapers cost-effectively

More technical issues regarding creation of projects.

preservation methods

Exploring how we all define "preservation" differently. A more knowledgeable copyright panel. If we do "team building" at the evening reception, find a way to capitalize on what comes out of it, instead of switching gears the next day.

Perhaps, there could be a web site with a collection of contacts and lessons learned.

No response

It would be useful to know more about news editorial and production systems. This is the foundation for figuring out the archiving piece.

So many papers are dying - we should talk about trying to save their archives before it is too late.

I suppose to try to form a systemic approach to try to save there, there were a lot of ideas, but nothing very organized.

We were very disappointed with the overall focus of the conference - that being the "monetization" of recent and historic newspaper content in order to save traditional newspaper publishing. We are a public archives with more than 135 years experience archiving and providing access to historic newspapers. We have devoted many public funds (both state and federal) to preserve and provide access to our state's newspapers and we are committed to providing free public access to those papers that were preserved with public funds. Without these efforts many of the papers "monetized" by the commercial vendors would not exist. The interests of non-academic, public institutions like ours were not adequately addressed or represented at the conference. Future conferences should make an effort to represent and address the challenges faced by such public institutions and programs, and emphasize the important role that commercial vendors, active newspapers, and academicians can play in advocating for continued public support of newspaper archives. We were hoping that the workshop on the second day would begin to address some of these issues and were sorely disappointed with the focus of the workshop. By restricting the workshop focus to the

"monetization" of newspaper content only, the discussions narrowly conceived and a waste of time. The only interesting discussions at the workshop occurred when participants tried to break away and address broader issues. The conference lacked a meaningful discussion of copyright issues and would be much improved by devoting more time to it.

More discussion between academic market and the news industry. We need to understand each other more.

Further development of Summit I recommendations. Can they actually be effective? -- How to convince your boss of the archive's importance and need for preservation. Perhaps a point-counterpoint from a publisher who understands the importance of the archives and one who sees the monetary potential of them. -- Who are our potential partners? Newspapers and media get advice on forming partnerships and the benefits of them. What are the most and least costly? -- Finding new ways to present historical content. Who are potential audiences, and what are the possibilities?

digital preservation of news from all media types - future cooperation between news publishers and libraries - legal deposit of born digital news such as practiced by Finland, UK, France, Norway, etc.

Who should be in attendance at Summit II who was not at Summit I?

More legal experts (copyright law).

More publishers of smaller newspapers.

Directors of some of the state projects funded by the U.S. Newspaper Program (California, Kansas, Pennsylvania, Kentucky, Virginia, and Illinois). They are all identified on the USNP web site maintained by NEH. (I should add that I am the director of the Illinois Newspaper Project.) I would also like to see more participation from CRL.

fewer vendors, more archivists, librarians, and digital preservation technologists

I don't think any important groups were missing entirely, but I think more representatives of large public libraries and state historical societies would be appropriate, as these institutions have both significant holdings of newspapers and a vested interest in their digitization

More library science students and faculty. I'm in that department, and it's really crucial that the students are aware of what's going on out there, as far as open access, big projects. I would like to have more opportunities for networking with the scholars and presenters, also.

historians, anthropologists, preservationists, archivists, librarians teacher have a point of view not based on how much more money it will cost or how to squeeze more profit out of what is by nature...commercially Copyright should expire in one (1)year. ,

Associated Press standards and technology gurus, IPTC, NAA

Perhaps James Simon from the CRL ICON project. http://icon.crl.edu/

No response

Representatives of the financial news community, from major organizations like Dow Jones/Factiva, Bloomberg. It was good to have the person from IBD at Summit I, but they are not the change leaders. And senior working journalists and editors.

I was hoping to see some of the people who microfilm newspapers in the surrounding states (such as Sheila Lee from Louisiana State University)

Brewster Kahle of the Internet Archive Ed Summers, Library of Congress Derek Willis, New York Times Maybe John Yemma, Christian Science Monitor editor and / or other news manager opinion leader - someone that the industry will actually listen to.

Also - just received a survey from an MLS candidate doing a project about this, see below Someone from the National Endowment for the Humanities - which routinely offers grants for local history digitization Dear Public Library Director, I am a graduate student at Southern Connecticut State University, and I am requesting your help in conducting research aimed at learning about "Digitizing Local, Community Newspaper Content in Small and Rural Libraries in New England: A Survey of Content Access to Patrons." I am asking library directors or qualified librarians in New England to participate in a survey. The survey results will be used to complete a study about the current state of local community newspaper digitization in small and rural public libraries. The results of this study will potentially be useful to all small and rural libraries with an interest in completing digitization projects of their local and unique collections, and specifically newspaper digitization, which can present unique challenges. If you are able to help me with this study, please complete the survey at https://www.surveymonkey.com/s/G77WFLB. Return of this survey indicates your consent to have your data used in this research. This survey will be open until the end of June, 2011. This survey is completely anonymous and the data will be compiled and kept in a file on my password protected computer and on a flash drive in a secure locked place for three years. Please note that the last question asks for contact information if you are willing to participate in an in-depth interview. This contact information will be separated from survey results and any information used from interviews will also be anonymous. No names will be published in the finished product. If you require a copy of my IRB application I will be

happy to e-mail you a copy. Please contact me at brikiatiss1@owls.southernct.edu with any questions or concerns. Thank you for your time and attention. Sincerely, Sylvie Brikiatis Dr. Yan Q. Liu – Project Advisor MLS Degree Candidate Southern Connecticut State University Southern Connecticut State University Liuy1@southernct.edu Brikiatiss1@owls.southernct.edu (203)-392-5763

A better representative of public newspaper archives and the issues they face in dealing with active newspapers and commercial vendors as well as those issues raised in #4.

more newspaper publishers. I'm not sure.

-- New York Times, always an industry leader. -- The History Channel, which now has an interesting magazine and membership concept. Is it working? -- Noted historians and authors, and how they work with archives.

NewsBank / Readex, someone from Bibliotheque Nationale de France, someone who represents digital preservation such at Martin Halbert, newspaper publisher with a pay wall such as Wall St. Journal or NY Times, someone from AP, a couple of people from news publishers with a different business model such as Google News, Small Town Papers, TownNews, etc.

About the author

Victoria McCargar is veteran journalist, archivist and digital curation consultant. Her 25-year career at the *Los Angeles Times* included planning and implementation of a number of newsroom computer systems, including digital asset management, pagination and archives. In the area of digital preservation she participated in the PREMIS (Preservation Metadata Implementation Strategies) project and was a researcher for the InterPARES consortium. Specializing in news assets, she has done research into long-lived news databases for the Center for Research Libraries. Consulting clients include the *Atlanta Journal-Constitution*, the *Columbia Missourian*, Associated Press, National Public Radio, the *Los Angeles Times*, Tribune Co., the University of Missouri Libraries, and the Annenberg Foundation. She is currently the archivist and special collections librarian at Mount St. Mary's College and teaches preservation management in the library and information science programs at San Jose State University and UCLA. She holds master's degrees in library science from UCLA and in science journalism from the University of Missouri-Columbia.