Collecting, Preserving, and Transforming the News –for Research and the Public
27-28 April 2017
Landshóbksafn Ísland—Háskólabókasafn (The National and University Library of Iceland), Reykjavik, Iceland

The Academic News Librarian—Patron Services to Support Text Mining

Debora Cheney
Assistant CIO for Assessment and User Engagement
Information Technology Services, University at Albany, SUNY, Albany, NY, USA.
dcheney@albany.edu

Abstract:

Much research and commentary focuses on the role of memory institutions and news organizations collecting, preserving, and transforming news for research and the public. These efforts increase the availability of news suitable for text mining. This change presents an opportunity for academic news librarians to provide services to researchers using news content which support the scholarly communication life-cycle related to this relatively new research methodology. Yet, little attention has been given to the nature of these services, the challenges researchers face, and the knowledge and skills news librarians and researchers will need. "The news" is different from other literary/book forms and today's news environment presents many challenges to researchers. Text mining is used across many disciplines, often by researchers who view the news text data as a source for information and patterns unrelated to the news itself and can benefit from the insights and support of news librarians. As subject experts on news resources, librarians can serve as consultants and educators helping researchers define "the news" they will use, by helping them negotiate challenges related to the library's collection, by locating news content which may extend well beyond the library's traditional news collections. This paper will elaborate on how the difference in news media creation, distribution, and readership impacts how researchers articulate their research question and outcomes. The challenges of locating news content are familiar to academic news librarians, but news text data mining presents additional challenges and opportunities for academic libraries to consider how they will support this new research methodology. The challenge is whether libraries are well-positioned to support such research or whether there is a disconnect between a library's preservation efforts and its patron services.

Keywords: newspapers, news media; patron services; academic libraries; text mining

Much research and commentary focuses on the role of memory institutions and news organizations in collecting, preserving, and transforming the news for research and the public. Certainly, these efforts will continue to increase the availability of a corpus of news text of high-quality and historical-depth regardless of whether it was born-text or born-digital. Less attention has focused on library services to support the use of news content for research, teaching, and learning.

The form and function of the news provides a great deal of opportunity for researchers who seek to use news content for their research, including text mining and other methods, for the "identification and exploration of interesting patterns"(Feldman & Sanger, 2007). It also presents a significant opportunity for academic libraries to provide services to researchers, faculty, and students using news content. These additional services will strengthen and extend the role libraries play/serve in supporting the use of news content through the entire the scholarly communications lifecycle including "generating, discovery, gathering, analyzing, translating, and repurposing (Davis-Kahl & Hensley, 2013, p. v).

News librarianship, a specialization based primarily in news organizations, has undergone significant challenges, changes in roles and responsibilities and threats to its existence. On the other hand, changing news production and distribution and new research methods of interest to many disciplines are creating a greater need for academic news librarians who are familiar with news content in all its formats, its production, readership, and preservation. These librarians will be able to develop comprehensive and well-developed services beyond collecting and preserving to support news text mining.

News Librarianship: Some Context

> We may not be on the frontline of journalism…but without news librarians
> beavering away in the background, arming reporters with facts and figures, the
> news would not make it into print—Katy Stoddard, Freelance researcher and
> writer, The Guardian, 2016 (Stoddard, 2016)

Although once primarily responsible for the news organization's clipping file (Hegg, 1991), today's news librarians, while fewer in number (Silverman, 2010), are receiving by lines on Pulitzer Prize-winning stories ("Episode 9: Caryn Baird on news librarianship," 2016) conducting in-depth research and analysis, and training journalists to become more effective researchers. Causing one Dean of Communications to argue: "I think journalists could learn a lot from hanging around with successful librarians"(Hare, 2016). In response to the changing environment in the news industry some news librarians are re-tooling and developing greater skill in programming and other computer-based research methods (Edds, 2017).

While the number of news librarians and news organization are declining, news librarianship is a growing need in academic libraries. The visibility and availability of news in everyday life results in a population that feels it is steeped in, and, even overloaded, with news—that appears to be always available and easily accessed to the extent that "news fatigue" has begun to set in (Ritter, 2008; Salzman, 2013). In addition, growing academic interest in cross-disciplinary research, use of news content for research unrelated to the news itself (Yzaguirre, Smit, & Warren, 2016), news as a first draft of history, a cultural record, and complement to a democratic society, as well as new digital-based research methods fuels an interest and growth in the use of news content within research, teaching, and learning.

Yet, academic news librarianship is a largely un-heralded expertise within librarianship. Certainly there are news librarians in academic libraries. However, their focus has varied

with the decade and the institution—from librarians who head libraries supporting the students and faculty in the journalism program; those who often work primarily in the periodicals or microfilm department where the bulk of the newspaper collection is housed; subject liaisons (often in media studies, communications or history) who also support the library's news collections; as well as preservation librarians and archivists with a particular interest and commitment to preserving newspapers.

Newspaper collections, like government documents collections, have always reflected libraries inherent role to preserve content and formats of historical and cultural importance. Libraries played crucial roles in preserving newspapers (in paper and later on microfilm) and large newspaper collections served as regional repositories of news content available to local researchers. For the most part, academic news librarianship has not been, but should be, considered a specialization similar to government documents which typically requires expertise in format, content, and subject matter.

While libraries and librarians differ, what is needed in many large academic libraries is strong and comprehensive subject expertise in newspapers and news content in all formats from born-print to born-digital. Many librarians receive little training in how news is created, published, preserved, and, how, now, news text has a continuing life as digital text separate from its origin on the newspaper page, on a website, in a blog, or connection to a twitter hash tag. Academic news librarianship requires: in-depth knowledge of the journalism profession, how news organizations work and how they distribute their news content; as well as detailed knowledge of what is and is not included in their library's collections from the earliest newspapers, international news content (Simon, 2015), and every format of each—including print, microfilm, and digital formats--as well as knowledge of the challenges and processes related to preserving news content.

The News: Some Context

"Wave after wave of digital innovation has introduced a new set of influences on the public's news habits. Social media, messaging apps, texts and email provide a constant stream of news from people we're close to as well as total strangers. News stories can now come piecemeal, as links or shares, putting less emphasis on the publisher. And, hyper levels of immediacy and mobility can create an expectation that the news will come to us whether we look for it or not"—The Modern News Consumer, Pew Research Center (Mitchell, Gottfried, Barthel, & Shearer, 2016)

Clearly, "the news" is significantly different from other literary and book forms. Faculty, students, and researchers face these fundamental differences when they begin a research project and consider whether using the news will be appropriate for their research question. These differences include patterns of distribution, readership, and the professional context of journalism and reporting

In fact, "the news" in our most recent past, consisted of newspapers and news programming by mainstream media companies. Yet, today's news environment (Project for Excellence in Journalism, 1998) is clustered around what is an ever-smaller amount of original news reporting and content (Pew Research Center's Journalism Project Staff, 2010) by fewer journalists working for these mainstream media companies. The growing presence of citizen journalists, the decline in the number of newspapers, and increase in and ongoing news commentary from nearly every perspective has changed the news environment dramatically. The news is less likely to be published in a print format, and more likely to be observed and appear on social media, then to be reported by a person trained in journalism. Michael Ferro, Chairman of Tribune Online Content, has asked: "When do social media

companies drop the word social and acknowledge officially they are media companies with the same editorial and ethical responsibility to provide real journalistic content to the communities they serve" (Warren, 2017)? The news is also interspersed with more and more commentary and punditry (Jurkowitz, 2013)--which is not, strictly speaking, news at all—although readers and researchers should begin to distinguish news reporting from opinion presented as news (Maines, 2013).

The news appears to be equally accessible in and to all parts of the world (although it is not), that all events are covered equally (although they are not), and that journalists are either, depending on your perspective, either equally biased or equally professional, in how they cover the news. Yet a great deal of social science and historical research finds that coverage of international and local news events demonstrates significant gaps in coverage and reporting related to specific news events (wars, police shootings, sexual assault, to name a few) (Meyers, 2004) is problematic in what is said and left unsaid. Even word choice and changes over time can present significant challenges for researchers seeking patterns in news text data. In fact, "the news" constitutes a very rough first draft of history often written from a very narrow or specific lens.

In addition, the form of the news also impacts whether it will be effective for some research questions: The 5Ws (who, what, when, where, and why), the lede, the use of significant amounts of repetitive content in the form of wire services stories, for example, but also in placing the context of a story (references to Columbine in all school shootings, for example)—all these, and more will impact the presence, frequency, and availability of news stories and its legitimacy as an appropriate source for news-based text mining. Whether "the news" has an impact on a reader's understanding of complex issues such as climate change (Schmidt, Ivanova, & Schäfer, 2013), on political events (Junqué de Fortuny, De Smedt, Martens, & Daelemans, 2012), or stock prices, (Li et al., 2014) for example, is extremely more complex than simply identifying digital news text to mine or to search. Researchers may need help understanding that news reporting may not always be the best sources to use as data (Franzosi, 1987).

These challenges are compounded when news text-based research(Gibbs & McKendrick, 2015) uses social media, such as *Twitter*, APIs and other webpage/data scraping technology (Norvag & Oyri, 2005) to obtain news data for text mining. Aside from the technical challenge, whether Twitter news is "the news" can be (and is being) debated (Wexler, 2014) and researched (Gabielkov, Ramachandran, Chaintreau, & Legout, 2016). Research analyzing the "geographic profiles" of Twitter compared with mainstream media "suggests that *Twitte*r is not simply a mirror of mainstream media, but rather has a distinct geographic profile, and that the differences between social and mainstream media geographic coverage deserve further exploration" (Leetaru, Wang, Cao, Padmanabhan, & Shook, 2013). In addition, Clifford Lynch recently suggested social media is part of the social record of news, but questioned whether it is journalism—but rather serves as an informal, social, observational and incomplete record of news events (Lynch, 2017). Certainly, it does not conform to traditional journalism standards for sourcing and verification, although it is used extensively by traditional media companies to generate interest and awareness of breaking news and generate readership (Swasy, 2017).

The significant changes in the availability and visibility of news content and significant preservation efforts have resulted in growing interest by faculty and students in the news and how it can be incorporated into their research, teaching, and learning. Even very recent political events, growth in fake news, and growing polarization of the news climate presents an opportunity for an enhanced but important role for academic news librarians.

The Academic News Librarian: New and Enhanced Roles

Evidence suggests the news content in many academic libraries is not used as heavily as it could be by researchers and students (Cheney, 2010; Feeney, 2014). The reasons for this have not been explored or fully understood. However, the impact of microform formats, the lack of indexing for many newspapers, and the lack of availability of many newspapers in some libraries likely impacted the use of newspapers for all but the most committed. As news has become more visible and accessible, even the daily newspaper collection has nearly disappeared in most academic libraries with many libraries believing researchers and students can use aggregator databases and Internet sources for this content—thus, believing there was less reason for the library to add news content to their collections. In this way of thinking the libraries' collection may be associated in the minds of many researchers, more with historical resources and preservation efforts—less so with today's news content. Libraries, also, may have come to believe there is no role for the library to play related to news content. However, nothing could be further from the truth.

However, access to news content in library collections is challenging, without even factoring in the wide range of news content that cannot be collected or preserved. Conway's survey of *The Library in the Life of the User* begins to suggest that newspaper access is considered less convenient and less helpful than other sources, such as an Internet search engine, electronic databases, and even library catalogs (Conway, 2015). A study of the University of Florida's Digital Newspaper Library suggests there are a variety of ways libraries could improve the user experience related to their digitized newspaper database interface (Reakes & Ochoa, 2009).

What is needed are more academic news librarians who can work closely with faculty, students, and their fellow librarians, to support the use of news content for research, teaching, and learning.   More discussion and awareness is needed in how libraries can support news-based research. Academic news librarians, serving as subject liaisons are needed to help "weave" together the threads of scholarly communications and information literacy (Davis-Kahl, Fishel, & Hensley, 2014) as they apply specifically to news-based research.

The Information Cycle: News and Media Literacy

"Whether you call it media literacy, news literacy, digital literacy, information literacy, it all deals with building critical thinking skills—Clark Bell, McCormick Foundation Journalism Program Director (Jolly, 2014)

Today noteworthy news events rarely move through a "progression of media coverage as depicted in an information cycle"("The Information Cycle," 2016)Rather, today's news environment is so complex researchers will first need to do some careful analysis of the news itself, to determine whether a news source is appropriate for their research question. Yet, a recent Stanford University study demonstrates that many students cannot distinguish between fake news and real news (Donald, 2016). David Roberts, a writer on climate change, describes media coverage as "tribal epistemology"—with readers judging information "not on conformity to common standards of evidence or correspondence …but rather on the values and goals…vouchsafed by tribal leaders" (Roberts, 2017) and Donald A. Barclay, *deputy university librarian at the University of California, Merced has* argued: "Today, the situation is far more nuanced. And not just because of the hyperpartisan noise of social media" (Barclay, 2017).

Faculty, students and researchers need to develop greater news and media literacy. These are crucial life-long skills for all researchers and students, generally, but also for journalism students. They are crucial for news-based research. As news consumers, many researchers have no in-depth knowledge about how the news has been, or currently is, created and distributed. Although readers and researchers may have clear biases toward or away from specific media companies or sources—researchers need to understand the difference between local newspaper and national news coverage. Research in local news can identify developing national problems (Cheney, 2015, p. 138), and, most recently, that the local news may have provided crucial coverage that affected the coverage of major Congressional legislation (Lieberman, 2017). All these skills and knowledge—can be taught by academic news librarians as collaborators and consultants supporting news based research.

Libraries can contribute to the discussions or research related to news and media literacy (Banks, 2016). To date, these discussions and efforts have largely been led by academics outside libraries, by organizations and non-profits. Academic news librarians could make significant contributions to this discussion focusing on areas which are rarely addressed, including developing a detailed understanding about the quality, availability, and appropriateness of specific news text for research, teaching and learning and the role the library can play in providing ongoing access to news content that follow important standards for news preservation.

News as Text—Data Literacy

Data literacy shares some distinguishing features with media literacy, where the use of tools to use and reuse content in ways not imagined by the content creator are a critical part of this literacy. Ownership and rights issues are core in both these, and librarians need to be able to address these at all levels of instruction (Association of College and Research Libraries. Working Group on Intersections of Scholarly Communication and Information Literacy, 2013, p. 11)

Academic librarians have always faced significant challenges related to locating newspapers for researchers—these challenges continue as formats, news creators, news aggregators and distributors continue to change and proliferate. Leetaru states "data collection is often regarded as the easiest step of the entire analytical process. Yet it is actually one of the most complex with the quality of the collection and preparation processes impacting very other stage of the project" (Leetaru, 2011, p. 7). Finding news text data suitable for text mining presents many challenges (Cheney, 2015).

Patterns found in text mining ultimately reflect all the strengths and weakness of the data itself, but will it be clear to today's researchers whether the data represents reliable and valid representations (Franzosi, 1987, p. 6) or whether the data was sufficiently complete (whatever that means), completely searchable and accurately reflected the content we consider to be news. This can present significant challenges for news content in aggregator databases, for example, whose content is often not thoroughly documented and may overlap significantly with other news sources or archives.

Although a great deal of effort can be spent locating news content, additional expertise in data literacy will be needed to ensure the news text data is appropriate for text mining. Data literacy requires "understanding how to find and evaluate data, the version of data being found and used, who is responsible for it, what standards are being used to preserve the data, how to cite the data, and ethics of data procurement. Data literacy is an area where the impact of external forces . . . point to a critical area of intersection between scholarly communication

and Information literacy" (Association of College and Research Libraries. Working Group on Intersections of Scholarly Communication and Information Literacy, 2013, pp. 10–11).

Today, there is no complete digital news archive of born-digital news—in the same way microfilm captures a newspaper page. Even what is included and not included in aggregator databases can vary from source to source and aggregator-to-aggregator. For digital archives of print newspapers, the quality of the underlying OCR, impacts the completeness and the appropriateness of the news text or source for text mining.

In fact, contrary to what many researchers might believe, Carner, et al found that news content in digital format are less likely to be preserved by news organizations (Carner, McCain, & Zarndt, 2014). News corporations themselves, face significant challenges ensuring they have an accurate archive of both their print-only, print/digital, and digital only content (Valkenburg & Sandhaus, 2016) and memory institutions are still seeking viable partnerships, projects, and funding to preserve born-digital news content in significant quantities for a significant number of titles to ensure this content is being preserved (Zwaard, 2017) The bottom line: news that is born digital, despite its apparent visibility and availability is frequently the most difficult news content for researchers to gain access to and use—simply because there is no open or affordable access to this content.

From the researcher's perspective it is not uncommon to approach a newspaper archives with the belief they are rich sources of data, soon to realize the value of the data is "not easily extracted from the newspaper archives. These archives are typically unstructured and organized by date rather than by them/subject or place. Metadata is often minimal and does not does not indicate which articles are important, which articles are outdated or duplicated, and which will be of long-term interest, particularly when the interest is driven by regional or personal interest" (Yzaguirre et al., 2016, p. 2). These are significant challenges to the researcher, and not all of them within the librarian's role to resolve, but academic news librarians may be able to recommend more appropriate sources, assist in developing a strategy for extracting the data using the news structure, or suggest effective technology that would allow the researcher to move forward.

Academic news librarians will work more extensively outside the scope of their collections assisting researchers with access to news content that lies outside traditional library resources, outside "old media" and by negotiating licensing restrictions, explaining copyright law and ensuring the ethical use and retention of this content. News content which is being used separate from its original form should be better understood as non-consumptive (Magier, 2012) and transformative use, which is disambiguated from the original form and function. Academic news librarians can explain these concepts and become advocates for embedding this understanding about the use of news-text data.

They will be able to promote and educate researchers in the role of the libraries' institutional repository in the lifecycle of their news-based research. Librarians are already visible advocates for broad use of content they have already paid for to support new research methodologies.

Emerging technology: Text mining and digital research methods

> As libraries continue to integrate more technology in frontend services and backend workflows, new positions will be created to meet individual libraries'' needs and existing positions will need to adapt in order to meet the expectations of our users and ourselves.(Radniecki, 2013, p. 4)

As some news librarians have done, academic news libraries may need to consider what their role will be in supporting digital research methodologies. In many libraries this service

is provided through the digital scholarship or digital humanities center within the library or university. However, the expertise of the academic news librarian should be available to support news-based digital research method, as news content is significantly different from other literary and media forms and requires the expertise described above to ensure the research question is appropriately stated to the form, function, and content of the news source(s) being used.

Evidence suggests a change is already underway within libraries that will allow academic news librarians to extend the scope and nature of the services they provide. Radniecki found it is still difficult to compile competencies for emerging technology libraries because these jobs are evolving within academic libraries (Radniecki, 2013). A recent Spec Kit on The Evolution of Library Liaison found liaison librarians are providing services related to scholarly impact and metrics, promoting the institutional repository, consulting on open access issues, e-research support and consulting on intellectual property issues, and new literacies education. "Examples of other services are data visualization support, GIS support, help with systematic reviews, text mining and promotion of open access journal development" (Miller & Pressley, 2015, pp. 14–15). Whether academic news librarians, like news librarian will need to re-tool and re-train to provide these skills depends on the a variety of factors, including the presence of other librarians with these skills in their library or university, the amount of collaboration with these digital/data research centers, and the amount of text mining research at the university. The challenge in most libraries is to ensure that this news text mining research is not separated from the academic news library and academic news librarian who possess particular knowledge and expertise in news content. There is an important role for libraries to provide this research support beyond the walls of their libraries and beyond traditional services

Conclusion

Academic news librarians are focused broadly on the collections and on providing access to news content, supporting preservation efforts, but also and on educating students and faculty about the role of news content in research, teaching, and learning—on developing life-long critical thinking skills. Use of the news text for research, teaching and learning requires, strong knowledge and literacy to support its use across the university curriculum, and to ensure researchers fully understand how their research question can (or cannot) be placed within the news context. The new news environment presents academic libraries have opportunities to help students and researchers both by integrating and supporting news-based research, generally, and text mining, specifically. As subject experts on news resources and news content, academic news librarians can serve as educators increasing knowledge and literacy related to news content, its form, function, and purpose; as consultants, collectors and negotiators extending beyond the library's traditional news collections by negotiating the challenges related to accessing, discovery, and using news content in compliance with licensing and copyright law; and providing technical expertise in a wide range of news-based research methods and tools.

References

Association of College and Research Libraries. Working Group on Intersections of Scholarly Communication and Information Literacy. (2013). *Intersections of scholarly communication and information literacy: Creating strategic collaborations for a changing academic environment*. Chicago: ACRL. Retrieved from http://www.acrl.ala.org/acrlinsider/archives/6970

Banks, M. (2016, December 27). Fighting fake news: How libraries can lead the way on media literacy. Retrieved April 3, 2017, from https://americanlibrariesmagazine.org/2016/12/27/fighting-fake-news/

Barclay, D. (2017, January 6). Column: Can librarians help solve the fake news problem? Retrieved March 30, 2017, from http://www.pbs.org/newshour/updates/column-can-librarians-help-solve-the-fake-news-problem/

Carner, D., McCain, E., & Zarndt, F. (2014). Missing links: The digital news preservation discontinuity. Presented at the IFLA World Library and Information Congress, Newspaper Section Satellite Meeting, August 13-15, 2014, Lyon, France. Retrieved from https://mospace.umsystem.edu/xmlui/handle/10355/43675

Cheney, D. (2010). Connecting the dots: How researchers use their library's news resources. In *Digital preservation and access to news and views* (pp. 68--89). Indira Gandhi National Centre for the Arts, New Delhi: IFLA Newspaper Section. Retrieved from http://www.ignca.nic.in/ifla2010/IFLA_PDF/IFLA_conf_papers.pdf.

Cheney, D. (2015). Big data, text mining and news content: Where is the big data? In *Contemporary research methods and data analytics in the news industry* (pp. 133–151). Hershey, PA: IGI Global. Retrieved from http://www.igi-global.com/chapter/big-data-text-mining-and-news-content/132911

Conway, L. S. (2015). *The Library in the Life of the User: Engaging with people Where they Live and learn*. Dublin, OH: OCLC Research. Retrieved from http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-library-in-life-of-user.pdf .

Davis-Kahl, S., Fishel, T., & Hensley, M. K. (2014). Weaving the threads scholarly communication and information literacy. *College & Research Libraries News*, *75*(September), 441–444.

Davis-Kahl, S., & Hensley, M. K. (Eds.). (2013). *Common ground at the nexus of information literacy and scholarly communication*. ACRL. Retrieved from http://www.alastore.ala.org/detail.aspx?ID=4383

Donald, B. (2016, November 22). Stanford researchers find students have trouble judging the credibility of information online. Retrieved March 31, 2017, from https://ed.stanford.edu/news/stanford-researchers-find-students-have-trouble-judging-credibility-information-online

Edds, C. (2017, February 20). The note I knew would come one day but I never wanted to write.

Episode 9: Caryn Baird on news librarianship. (2016, September 1). [.mp3]. Retrieved from http://beyondthestacks.info/index.php/page/2/

Feeney, M. (2014). Who uses international newspapers? A Content analysis of dissertations and theses. Presented at the IFLA International Newspapers Conference 2014, Salt Lake City, UT. Retrieved from https://www.ifla.org/files/assets/newspapers/SLC/2014_ifla_slc_feeney_-_who_uses_international_newspapers_.pdf

Feldman, R., & Sanger, J. (2007). *The text mining handbook*. Cambridge University Press. Retrieved from http://www.roelsbeestenboel.nl/text.pdf

Franzosi, R. (1987). The Press as a source of socio-historical data: Issues in the methodology of data collection from newspapers. *Historical Methods*, *20*(Winter), 5–16. https://doi.org/10.1080/01615440.1987.10594173

Gabielkov, M., Ramachandran, A., Chaintreau, A., & Legout, A. (2016). Social clicks: What and who gets read on Twitter? Presented at the ACM SIGMETRICS / IFIP Performance 2016. Retrieved from https://hal.inria.fr/hal-01281190/document

Gibbs, W., & McKendrick, J. (Eds.). (2015). *Contemporary research methods and data analytics in the news industry*. Hershey, PA: IGI Global. Retrieved from https://www.amazon.com/Contemporary-Research-Methods-Analytics-Industry/dp/1466685808

Hare, K. (2016, April 22). 5 Tips for journalists from a Pulitzer-winning researcher. Retrieved March 30, 2017, from http://www.poynter.org/2016/5-tips-for-journalists-from-a-pulitzer-winning-researcher/407942/

Hegg, J. L. (1991). Small newspaper libraries: The Libraries that time (and Automation) passed by. *Special Libraries*, *82*(Fall), 274–281.

Jolly, J. (2014, September 4). News literacy vs. media literacy. Retrieved March 30, 2017, from http://www.cjr.org/news_literacy/news_literacy_vs_media_literac.php

Junqué de Fortuny, E., De Smedt, T., Martens, D., & Daelemans, W. (2012). Media coverage in times of political crisis: A text mining approach. *Expert Systems with Applications*, *39*(14), 11616–11622. https://doi.org/10.1016/j.eswa.2012.04.013

Jurkowitz, M. (2013, June 5). Is MSNBC the place for opinion? Retrieved from http://www.pewresearch.org/fact-tank/2013/06/05/is-msnbc-the-place-for-opinion/

Leetaru, K. (2011). *Data mining methods for the content analyst: An Introduction to the computational analysis of content*. New York: Routledge.

Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013, April 22). Mapping the global Twitter heartbeat: The Geography of Twitter. *First Monday*, *18*(5). Retrieved from http://journals.uic.edu/ojs/index.php/fm/article/view/4366

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., & Chen, Y. (2014). The effect of news and public mood on stock movements. *Information Sciences*, *278*, 826–840. https://doi.org/10.1016/j.ins.2014.03.096

Lieberman, T. (2017, March 24). How local news sounded the alarm over the GOP's defeated health plan. Retrieved April 3, 2017, from http://www.cjr.org/united_states_project/how-local-news-sounded-the-alarm-over-the-gops-defeated-health-plan.php

Lynch, C. (2017, January 26). Clifford Lynch: Born-digital news preservation in perspective. Retrieved April 3, 2017, from https://www.rjionline.org/stories/clifford-lynch-born-digital-news-preservation-in-perspective

Magier, D. (2012). Recent trends in text mining and library services: Research library perspective. Presented at the Text-Mining Opportunities and Challenges. Retrieved from www.crl.edu/sites/default/files/follow_up_material/Text-Mining.pdf

Maines, P. (2013, August 15). Opinion journalism vs. objective news reporting. Retrieved from http://www.mediacompolicy.org/2013/08/15/opinion-journalism-vs-objective-news-reporting/

Meyers, M. (2004). African American women and violence: gender, race, and class in the news. *Critical Studies in Media Communication*, *21*(2), 95–118. https://doi.org/10.1080/07393180410001688029

Miller, R. K., & Pressley, L. (2015). *Evolution of library liaisons* (Spec Kit No. 349). Washington DC: Association of Research Libraries. Retrieved from http://publications.arl.org/Evolution-Library-Liaisons-SPEC-Kit-349/

Mitchell, A., Gottfried, J., Barthel, M., & Shearer, E. (2016). *The Modern news consumer: News attitudes and practices in the digital era* (Numbers, facts and trends shaping the world). Washington, D.C.: Pew Research Center. Retrieved from http://www.journalism.org/2016/07/07/the-modern-news-consumer/

Norvag, K., & Oyri, R. (2005). News item extraction for text mining in web newspapers. In *International Workshop on Challenges in Web Information Retrieval and Integration* (pp. 195–204). https://doi.org/10.1109/WIRI.2005.27

Pew Research Center's Journalism Project Staff. (2010). *How news happens: A Study of the news ecosystem of one American city* (Project for Excellence in Journalism). Washington, D.C.: Pew Research Journalism Project. Retrieved from http://www.journalism.org/2010/01/11/how-news-happens/

Project for Excellence in Journalism. (1998, March 6). Changing definitions of news. Retrieved January 3, 2017, from http://www.journalism.org/1998/03/06/changing-definitions-of-news/

Radniecki, T. (2013). Study on emerging technologies librarians: how a new library position and its competencies are evolving to meet the technology and information needs of libraries and their patrons. Presented at the IFLA WLIC 2013, Singapore. Retrieved from http://library.ifla.org/134/

Reakes, P., & Ochoa, M. (2009). Non-Commercial digital newspaper libraries: considering usability. *Internet Reference Services Quarterly*, *14*(3/4), 92–113. https://doi.org/10.1080/10875300903336357

Ritter, K. (2008). Study tracks news "fatigue" among younger readers. *Editor & Publisher*, (June 2).

Roberts, D. (2017, March 22). Donald Trump and the Rise of Tribal Epistemology. *Vox*. Retrieved from http://www.vox.com/policy-and-politics/2017/3/22/14762030/donald-trump-tribal-epistemology

Salzman, M. (2013, January 18). 2013: The Year fatigue set in. Retrieved from http://www.huffingtonpost.com/marian-salzman/fatigue_b_2471860.html

Schmidt, A., Ivanova, A., & Schäfer, M. S. (2013). Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries. *Global Environmental Change*, *23*(5), 1233–1248. https://doi.org/10.1016/j.gloenvcha.2013.07.020

Silverman, N. (2010, January 29). Endangered Species: News librarians are a dying breed. Retrieved March 30, 2017, from http://www.cjr.org/behind_the_news/endangered_species.php

Simon, J. (2015). Access to Print, Digitized, and Born-Digitial Newspapers from Africa: The North American/Global Conundrum. Presented at the IFLA WLIC 2015, Cape Town, South Africa. Retrieved from http://library.ifla.org/1269/

Swasy, A. (2017, March 22). I studied how journalists used Twitter for two years. Here's what I learned. Retrieved March 30, 2017, from https://www.poynter.org/2017/i-studied-how-journalists-used-twitter-for-two-years-heres-what-i-learned/453162/

The Information Cycle. (2016, April 18). Retrieved March 31, 2017, from http://www.library.illinois.edu/ugl/howdoi/informationcycle.html

Valkenburg, S., & Sandhaus, E. (2016, July 26). The Future of the past: Modernizing the New York Times archive. Retrieved April 3, 2017, from https://open.blogs.nytimes.com/2016/07/26/the-future-of-the-past-modernizing-the-new-york-times-archive/

Warren, J. (2017, January 3). Here are the biggest media mysteries of 2017. Retrieved April 5, 2017, from http://www.poynter.org/2017/here-are-the-biggest-media-mysteries-of-2017/443892/

Wexler, A. (2014, February 12). Twitter is the 21st century newspaper. Retrieved April 5, 2017, from http://www.huffingtonpost.com/adam-wexler/twitter-is-the-21st-centu_b_4774976.html

Yzaguirre, A., Smit, M., & Warren, R. (2016). Newspaper archives + text mining = rich sources of historical geo-spatial data. In *IOP Conference Series: Earth and Environmental Science* (Vol. 34, p. 012043). Bristol, UK: IOP Publishing. https://doi.org/10.1088/1755-1315/34/1/012043

Zwaard, K. (2017, January 26). Kate Zwaard: Technology and community--Why we need partners, collaborators, and friends. Retrieved April 3, 2017, from https://www.rjionline.org/stories/kate-zwaard-technology-and-community-why-we-need-partners-collaborators-and

Appendix:

Academic News Librarian should have a strong understanding of the following:

- News is no longer an "old media" product
    - The "definition" of news is evolving—no longer means words on paper and increasingly encompasses commentary and punditry
    - "Social media" are now "media companies", without the same editorial and ethical responsibility to provide real journalistic content to the communities they serve."
    - The impact and readership of the news is affected by linking and forwarding behaviours used by individuals and media companies--not on your subscription or the location of your doorstep
    - News Is heard and seen more than ever before, however, readers may be "casual" or purposeful in how they obtain their news
    - Original news reporting is declining, much news is the same news duplicated in multiple sources
- As 'the news" is no longer primarily an "old media" product, but library collections are predominately "old media" available in aggregator news database –this will likely change over time —these challenges continue as formats, news creators, news aggregators and distributors continue to change and proliferate;
- News is global, but readership is not
    - but regions of the world still generate news print
- All news is local, but coverage is not equal for all local news events—especially as multiple newspapers covering a market/area as well as the decline in number of small, local/regional newspapers has occurred
    - However local media may have significant impact on local readers on some topics
- Of the journalism profession, professional ethics, and how the press operates
- The form function of the news including how traditional news stories are structured ad writing style (5Ws, lede paragraph) and how this provides a structure that may impact word presence (for example, the Columbine shooting may be mentioned in many more recent stories about school shootings)
- There is no "complete" digital news archive—in the same sense as preserving a newspaper page, even, or especially, news aggregator databases (depending on your perspective) are not complete
- Digital archives, especially when they provide open or low cost access, increase use of those newspapers and news content, but the quality of the digitization (OCR, for example) and metadata must be carefully evaluated before a text mining project is viable using such sources
- Access to news is typically limited by copyright and licensing restrictions and access to content in aggregator databases for large scale text mining projects without special permissions and access. This chokes the use of this content and requires librarians to work closely with these aggregators to obtain access to news content.

Patron Services Academic News Librarians can provide:

- Incorporate news literacy into the library's information literacy program and use of the Information literacy guidelines

- Educate faculty and students in the appropriate use of copyrighted and licensed news resources for research and teaching;
- Work closely with student journalists to ensure they have the appropriate research skills but are also aware of the important roles for librarians in supporting journalism and information gathering to support journalism
- Become advocates for broader use of content they have already paid for to support new research methodologies—and awareness of the use of news content disaggregated from its original form;
- Develop skill or knowledge of text mining technology and software to support researchers obtain news text data;
- Promote and educate researchers who are using news content in the role of the Libraries' institutional repository in the lifecycle of their scholarship