Introduction

THE MAKING OF THE DIGITAL PROVOCATEUR: HOW DATA, INTERACTION, AND DESIGN DRIVE TODAY’S AND TOMORROW’S CONTENT DECISIONS

For anyone living in the United States in the 1960s, the evening news was a defining moment toward the end of the day. Family members dutifully congregated by the television to watch and listen to luminaries such as Walter Cronkite, Harry Reasoner, and David Brinkley speak about important events in places that often seemed far away and foreign. The anchorman’s presence and auditory imparted the seriousness and importance of world affairs, implying that everyone, even young children, should devote undivided attention. During commercial breaks and after the programs, people discussed the news. In families, children listened or questioned, “What is Vietnam?”, “Who is JFK?,” or “What do we mean by civil rights?” Many of those conversations and the questioning carried on for days, and ultimately had a profound impact on national policies. The most striking example of the impact of broadcast television news on the national psyche was CBS anchor Walter Cronkite’s editorial following the communists’ February 1968 Tet Offensive in Vietnam, in which he opined that a U.S. victory may be out of reach in Vietnam, and that a negotiated peace may be the only resolution to the conflict (Folkenflik, 2009). With the war and many other issues broadcasted over the evening news, people took positions, they extrapolated, expressed their views and ideas, and challenged others’ views. For many people, the nightly news experience was communal, thought providing, engaging, controversial, and it engendered much discussion.

In times past, to access this world of news luminaries, one turned or pushed a switch on the television, chose a broadcast station from a limited number of channels (in many cases there were only three), and occasionally adjusted the “rabbit ears” antenna on top of the television to enhance reception. Aside from intermittent antenna adjustments, viewers’ physical interactions with the television were discontinuous. People sat back, watched, and listened as news organizations delivered information on the events of the day to them. While television technologies seemed unwieldy and archaic by today’s standards, what transpired in living rooms, restaurants, shops, and meeting places around the world was often a unique and captivating experience that, at times, carried over to the days subsequent to a news broadcasts. And yet, aside from what existed in public memory and dialogue, the broadcast itself was intangible and ephemeral. The news organization provided a service and people watch and listened, with little information exchanged or interaction taking place between the two.

Characteristics of these early media delivery forms (i.e., television, radio, newspapers, magazines, journals) embody and influence how online news is delivered to people and how people use and interact with today’s media technologies. In the following section, the authors examine attributes of delivery media formats, specifically the amalgamation or separation of news and medium, and the relevance
of human interaction both with the medium and the news organization. In the authors’ view, an initial overview of these factors is important because they are foundational to and have implications for digital news media, user experience, and the study of big data and analytics.

**BACKGROUND**

In the 1960s, the television was a relatively new delivery format for the average news consumer. Television (and radio) sets were content neutral; they could receive an entertainment as well as news broadcast, for instance. News organizations were not involved in their design, which was left to product designers and engineers whose primary professional interest was not news. Television and radio were electrical devices with electromechanical parts with which people had to interface to get news. There were electromechanical controls – such as volume knobs – that enabled people to get news and control its attributes. Unlike newspapers, the physical properties of television and radio made for the partitioning of the news and the physical medium used to deliver it. To get news, people had to adjust to a new layer of complexity and learned how to operate the controls.

Despite the separation, someone who owned a television or radio starting in the 1950 and 1960s could look inside to see and feel the dials and other physical components, and if nothing else to make a mental connection what controls looked like and how they functioned. This made it easier to understand how the controls worked. In years to come, news organizations would find this separation between news and medium even greater as they developed digital interfaces and new skeuomorphic modes of interaction, references to an earlier medium used as a metaphor (Gross, Bardzell, & Bardzell, 2014), so people could access, control, and traverse news in virtual spaces.

With digital interfaces, no such tangible connection exists between what controls look like and how they physically operate. Users’ access to news depends more on how well the software or designer represents the functionality of the delivery medium at the interface. In fact, because that same physical connection does not exist in digital interfaces, designers use concepts from previous familiar media to help people understand how to operate the interface. For instance, most people who use news websites are familiar with the video *Play* icon (large arrow pointing to the right) or an audio speaker icon to control sound volume. These icons have no real connection to how the digital interface actually works behind the scenes. Touching the *Play* icon on a digital interface, for example, does not start a magnetic tape moving in a rightward direction. However, because people are familiar with them or have learned about them from older technologies, interface designers use these concepts (e.g., large arrow point rightward) to help people to more easily understand the interface and how it works, which ultimately encourages people to return to a website.

This has significant implications for the work investment needed to design news interfaces (and positive user experiences) so that people understand them and their functionality. In some cases today, entire departments with teams of individuals are responsible for building digital interface controls so that news consumers can access the news and have a positive user experience.

As was the case with early television and radio, newspapers, magazines and journals brought news and information to people, delivered to homes, or purchased at newsstands. The periodical afforded a highly individualized experience, unlike television or radio that provided both individual and group experiences. The paper was a tangible material onto which a printer printed words and images. People interacted with it physically and there was more continuous physical interaction with it compared to a
television or radio. The newspaper, magazine or journal was malleable, portable, and disposable. The news and content was inseparable from the physical paper on which it was printed and there was a direct connection between what a reader did with the paper and how it impacted the paper. When the reader folded the paper, he or she could see and feel the direct result of his or her actions on the newspaper. Moreover, the newspaper, as the title suggests, was a medium devoted to news. People used it for the primary purpose of getting the latest information on events within their communities, states, nations, or across the world. It was rarely used for assorted purposes such as the reading of a biography or works of fiction.

The conception of newspapers, magazines or journals encompasses the paper material and the printed word. These publications were not content neutral. Moreover, periodicals had been following an established design formula for hundreds of years, with relatively minor modifications, so people thought little about how the design of the publications themselves impacted how easy or difficult it was to get news and information. With newspapers, major stories were printed with large-typeface heads “above the fold,” so they would be prominently displayed to catch the attention of readers. Magazines and tabloids highlighted major features prominently on their cover pages. People knew how to get information from the printed paper, as it was a familiar medium. And yet, while there may have been more direct and continuous interaction with periodicals (people touch them more than they did a television or radio), there was in this experience little exchanged or interaction taking place between publishing organizations and consumers.

Due to digital technologies, the Internet, and mobile connectivity, major changes are taking place today regarding how people get news. People interact more with various types of devices to get news and information. During these interactions, the content and physical medium are detached and mediated by a digital (virtual) interface, which often adds complexity to interactions and challenges to creating positive user experiences. A digital device is content neutral, as are television and radio. However, interaction is more regular and continuous - people touch it more for longer periods of time, not unlike a newspaper. Those points of contact (the interaction) at the interface and the technical capacity to record them have significant implications for journalists, user experience designers, editorial professionals and publishers seeking to interact with and better serve audiences by employing big data and analytics. They open up innumerable opportunities for ongoing interaction between publishing organizations and information consumers.

**INTERACTION, BIG DATA, AND ANALYTICS**

At a most fundamental level, technological advancement has incited more people than ever before to pick up and engage with technology. Their engagement with digital devices is more frequent and sustained over longer periods of time – more people are touching, speaking to, moving, looking at, clicking, gesturing, and interacting with digital technologies. At the same time, technologies now have the capacity to chronicle interaction events aggregated as data that can be studied. Streams of interaction data, despite how voluminous, can be amassed and analyzed regardless of the mode of interaction, the duration, frequency, or geographic location in which the interaction event occurred. As Chen, Chiang, and Storey (2012) indicate, at the beginning of the twenty-first century, Internet and web technologies (Web 1.0-type systems) presented unparalleled possibilities for data collection and measurement as businesses and organizations began interacting with customers one-to-one online. They further point out
that user-generated content aggregated with Web 2.0-type systems (e.g., social media) made possible
the compilation, measurement, and analysis of additional, novel types of data associated with textual
information and less structured content and these data hold valuable insights into customer opinions and
behavior; therefore aiding researchers in attending to and understanding the views and habits of their
target audiences. This is important because due to the growth of digital technologies, mobile connectivity,
the escalating effects of social media, and the overall increased usage of technology makes for a more
informed and discerning consumer (Pain & Lively, 2013). If organizations ignore consumers’ opinions
and behaviors, consumers will seek comparable good and services elsewhere.

WHAT IS BIG DATA AND ANALYTICS?

The Internet has made information throughout the world and associated media accessible to any person
on the planet with the touch of a button. Twenty-four hours a day, seven days a week, human connect-
ivity through mobile and desktop devices is unparalleled. Unlike any other period in human history,
people have the information of the world instantly accessible to them. Exponential growth of Internet,
technology-enabled connectivity, and increased accessibility to information engender new opportuni-
ties for observing human behavior, data collection, and measurement (Google, 2014).

Sensor-equipped devices, mobile technologies, the Internet of Things, “the vast repository that we
call the Web, user-generated content, social media, data generated and consumed on mobile platforms,
and data from enterprise systems” (Goes, 2014, p. iv) allow for monitoring human behavior, location,
emotion, and temporal patterns, among other things. Analyst, researchers, businesses, and organiza-
tions can now inquire about, explain and potentially predict human behavior patterns and trends using data
compiled from sensor-equipped devices and the Internet of Things (Goes, 2014). As Chen, Chiang, and
Storey (2012, p.1168) indicate,

Other sensor-based Internet-enabled devices equipped with radio-frequency identification RFID,
barcodes, and radio tags (the “Internet of Things”) are opening up exciting new streams of innovative
applications. The ability of such mobile and Internet-enabled devices to support highly mobile, location-
aware, person-centered, and context-relevant operations and transactions will continue to offer unique
research challenges and opportunities throughout the 2010s. Mobile interface, visualization, and HCI
(human–computer interaction) design are also promising research areas. Although the coming of the
Web 3.0 (mobile and sensor-based) era seems certain, the underlying mobile analytics and location and
context-aware techniques for collecting, processing, analyzing and visualizing such large scale and fluid
mobile and sensor data are still unknown.

Ninety percent of the world’s data today have been produced in the last two years from sources such
as sensors used to collect and model weather conditions, social media, digital media, online purchases,
and Global Positioning Signals (GPS) (IBM, 2014a). The sheer volume of data is a key factor instigat-
ing the title Big Data. There is a transformation in the scale of data and scope of the possible sources
as well as means for managing those sources (Taylor, Schroeder, & Meyer, 2014). According to Goes
(2014, p. vii), “Big data is about massive amounts of observational data, of different types, supporting
different types of decisions and decision time frames.”
Introduction

Big data or the high volume data streams (often in real-time) being collected today are distinctive because of the following four “V”s (see Chen, Chiang, & Storey, 2012; Waller & Fawcett, 2013; Goes, 2014; Taylor, Schroeder, & Meyer, 2014): (a) the *volume* or scale of data – volume can mean more data and/or data are captured in more detail, or it can also be the large amounts of sentiment data that accumulate from social media, or opinion commentaries (Waller & Fawcett, 2013); (b) the *velocity* or real-time or almost real-time analysis of data that can be achieved, the big data world is moving toward real-time or close to real-time decision making (Goes, 2014, p. vii); (c) *variety* of data – data are being collected with various technologies (e.g., sensor-equipped devices, mobile devices) and from a variety of sources, (e.g., social media, tweets, identification, emotion detection, eye tracking, etc.); and (d) *veracity* or ensuring the quality of data because often the quality of captured data vary and analysis depends on the accuracy of source data. Some contend that another “V” should be *value*, that having access to large amounts of data is not useful unless they can be used to add value to the organization.

Einav and Levin (2014) point out three important characteristics of big data. First, data sources are frequently real-time accessible, which is advantageous for making observations about present conditions or predicting the present, commonly known as *nowcasting*. A second characteristic is data volume, as mentioned earlier. Among other things, large data sets help “…resolves the statistical problem of limited observations and makes analysis more powerful and potentially more accurate, while their granularity increases their power in terms of understanding individual actions” (Taylor, Schroeder, & Meyer, 2014, p. 2). Third, because the breadth of data sources and information has expanded, data are less structured. Now researchers have the capability to identify, observe, and measure an almost limitless collection of distinct behavioral attributes. Additionally, data sources enable the collection of data relating to certain attributes of human behavior (e.g., social relations, sentiment, geo-location, etc.) that were previously problematical to observe.

Analytics is or should be linked to decision making (Goes, 2014) based on the data collected. It is a domain of study that has materialized as a dynamic research area (Chen, Chiang, & Storey, 2012) that offers business and organizations opportunities for continual improvement using the aforementioned large data sets. Digital analytics employs sophisticated methodologies to these large and distinct data (big data) sets; some of those methods include predictive analytics, data mining, text mining, web analytics, integrated statistics, opinion mining, social network analysis, spatial-temporal analysis, visualization, and summarization tools (Zhang, Chitkushev & Brusic, 2014). Kaushik (2010, p. 5) defines digital analytics as encompassing the following three elements:

1. *The analysis of qualitative and quantitative data from your website and the competition, (2) to drive a continual improvement of the online experience that your customers, and potential customers have, (3) which translates into your desired outcomes (online and offline).*

Because of the ubiquity of digital technologies, people in today’s society continuously interact with digital interfaces that mediate the exchange between human action and computing. At these points of contact or interfaces large amounts of data can and are being collected and these data (and related analytics) are shaping, among other things, publishing, news services, the news interfaces, the user experience, how news is gather and received, who participates in news making and news distribution.
THE NEXT MEDIA FRONTIER: DETECTING, RECORDING HUMAN-TECHNOLOGY INTERACTIONS

Consider how far along digital media has come, with technology widely and readily available today, to interact with audiences, in real-time, and understand and accommodate their informational requirements. Data streaming in from websites and devices provide a comprehensive picture of when, where and how audience members are consuming information. At the same time, consumers may be contributing responses and feedback in a passive way that does not require active engagement, if they choose not to do so.

The data now available from news and information consumers can determine what parts of a screen are getting the most attention, and even electronically detect the reader’s interest in the topics being delivered. The technology exists that can provide instantaneous services and feedback as consumers move through the day. As renowned science author Arthur C. Clarke once wrote, “Any sufficiently advanced technology is indistinguishable from magic.” This is certainly the case with the new generation of media technologies that is emerging, enabling consumers to receive, on their screens in front of them, the exact information they need, tailored to their reading habits and even to their moods. Bearing in mind Clarke’s statement, it would have been incomprehensible (magic) for most news consumers in the 1960s to imagine many of today’s innovations (existing or emerging) that have implications for big data collection. To provide a measure of historical context, the authors examine a few relevant examples that might have been perceived as miraculous just 50 years ago.

Few people in the 1960s could have envisioned that one day the screen they were watching might have electrical sensors connected to it that monitor eye movements, determining the locations on the screen where viewers’ fixations to process visual information or their saccades, the brief periods between eye fixations when perception is attenuated (Palmer, 1999). The data from these sensors could tell news organizations and advertisers what viewers found interesting, what caught their visual attention and, perhaps more importantly, what they did not see. Additionally, the sensors could help determine if viewers left the room during commercials, how much time they fixated on the screen, where they fixated the most, was the time spent gazing at the screen continuous or discontinuous, and the visual scan paths or the ways their eye traversed the screen. As people watched a broadcast on the presidential elections, for example, a switch on the device might enable people in living rooms nationwide to engage in conversation. The audio from these conversations or text they typed on a keyboard could be analyzed on the fly, to determine sentiment. Geographical maps with overlays (heat-maps) enable viewers to ascertain the tenor of discussions as well as the locations of contentious conversations so they could avoid them or join in.

This technology can be applied to the periodical space as well. For most news consumers in the 1960s, it would have been inconceivable to think that when they brought a newspaper or magazine at the newsstand, the publication could record their location, the time of day as well as their reading habits based on the hundred other times they brought periodicals. The publication would know by the location and time of day that the consumer is traveling to the office by public transportation – to most consumers just 50 years ago, that would certainly be magic. Consequently, the publication presents itself in a smaller viewport so the reader does not have to make adjustments to prevent from interfering with people in adjacent seats. The publication also knows that on the way to work, the news consumer reads major news stories of the day and the financial news and so it displays those sections in its viewport. Most news consumers couldn’t have imagined these things in the 1960s. Continuing this example a bit further, when the consumer travels home, he or she likes to read the world news and comics. Because the publication knows the consumer’s location and the direction he or she is traveling, world news stories and comics display or
are readily accessible on the viewport. On the way home, the consumer touches a travel advertisement on the display and then begins reading it and several other travel advertisements. As he or she exits the transportation terminal, related travel advertisements display on electronic billboards in the terminal. The publication detects that the individual may not have finished reading a particular news story. When he or she returns home and re-opens the newspaper or magazine, the point at which he or she left off is prominently displayed in the viewport. Alternatively, if he or she turns on the television or radio that story is listed as an option so he or she may continue learning about a particular topic or news event.

Fifty years ago, the news consumer could not envision that if he or she felt passionate about particular topics, he or she could write comments about them using the interface. Those comments get added to millions of other comments from people worldwide, enabling the individual to understand these topics from unique and varying vantage points and cultural perspectives. The commentary can be depicted graphically in maps and charts showing, among other things, the tenor of conversations, who is central to debates, the group formations, group splintering, homogeneity, and activity. Because people are discussing and exchanging information on topics, they extend the life of many stories. Accordingly, people have recording devices that enable them to record events in high-quality multimedia that can be shared with others and with publishing organizations. Thus, people, in conjunction with media organizations, actively engage in spreading news and content worldwide.

Finally, in concluding this example, imagine what consumers in the 1960s would have thought if the publication (device) came equipped with electrodermal response sensors to measure reactions. It can ascertain when the person is excited or upset, and thus may change background color and type size, the amount of content displayed, or it may simplify the interface to better suit the person’s state. If the device can determine that the person is in an anxious state, it might display or include in the interface advertisements for relaxing vacations or techniques to relieve tension. The device might also display a reminder screen so the individual can remember to schedule an appointment with a physician – and perhaps, if the individual prefers, this stream of electrodermal response data can be sent to his or her physician for monitoring.

The aforementioned applications may sound futuristic – they were certainly futuristic (magic) to news consumers 50 years ago – but most actually exist today because of advances in technology that facilitated developments in areas such as graphical user interfaces (GUI), personal computing, networking and information sciences, Web, and mobile technologies. From a human-technology interaction perspective, people’s increased and continued interaction with digital technologies and the technical capacity to chronicle those interaction events are requisite in the evolution of what is known today as big data and analytics. They also change many facets of the news and publishing industries and the ways information consumers interact and engage with them.

**THE MESSAGE IS THE PLATFORM**

For decades, the media – regardless if it was print or electronic – was primarily a one-way interaction. News and information was collected, selected, edited, packaged up and delivered to consumers, and consideration of its ultimate value to audiences, or exactly how it was being received, was only given on a high-level basis (see Figure 1). The acceptance and selection of content was only loosely tied to overall subscription sales of periodicals or journals. As long as these sales were profitable, it was generally assumed that audiences’ needs were being met.
Introduction

Up to even a few years ago, this one-way delivery of content was always regarded as the go-forward model of publishing. The interactivity of television and radio was limited to consumers selecting a station, and either remaining with that station, or tuning to another. Content adoption was based on the selection of samples, collected and reported to networks by specialized third-party firms such as A.C. Nielsen. This was also the primary form of feedback on consumer content preferences available to networks.

For print publications such as newspapers, magazines and journals, feedback on consumer content preferences was even more generalized. Subscription and newsstand sales on daily, weekly or month-to-month basis gave editors and journalists a very general picture of how well their material was being accepted. Publishers also conduct market research surveys and focus groups to develop a better picture of their audiences.

Even when the future of media was considered, it usually followed the traditional one-way delivery model. For example, a number of futurists predicted that newspapers would ultimately be delivered in the form of some type of electronic facsimile device to homes (Novak, 2013). There were predictions that content would be tailored to users’ preferences, but in a very broad sense. It was also assumed that traditional existing media outlets would be the providers of this information.

With the capability to chronicle human behavior (interactions), news and publishing organizations began investing in real-time data and associated analytics to better understand consumer behavior and to assess their own effectiveness. Nobody foresaw the rise of big digital, content-sharing platforms that would serve as disruptive news and information platforms. Social media sites such as Facebook, Twitter, Yahoo! and Google are playing prominent roles in information delivery, often serving as alternatives to traditional media – effectively competing for readership numbers as well as advertising dollars.

A survey of 5,173 adults conducted by The Pew Research Center in collaboration with the John S. and James L. Knight Foundation found that about half of both Facebook and Twitter users get news on those sites, while one-fifth of YouTube use the site for that purpose (Holcomb, Gottfried, & Mitchell, 2013). It is also reported that Facebook delivers almost a quarter of all digital advertising views in the U.S., and Google takes in almost half the dollars spent on online ads through its search advertising (Grueskin, Seave, & Graves, 2011).

The value of social media platforms is not lost on traditional media companies. Many are leveraging social media sites to amplify their content, or to better interact with their audiences. Media organizations, for example, are setting up “fan pages” on Facebook, along with Twitter feeds, in which they encourage readers to share links. Media organizations also include social sharing tools on their websites, including Facebook, LinkedIn, Twitter, Digg and Reddit.

Online platform providers continue to seek ways to play more prominent roles in the media markets. At the time of the first authorship of this book, Snapchat, the mobile-based instantaneous video sharing service, announced that it was partnering with media companies such as CNN, ESPN, Vive and Warner Music to “skim” selections of their multimedia content and sell ads against the videos, photos and articles, and share the revenue with the publishers (Kafka, 2015).

The advantage that many of these new platforms provide is greater transparency and enhanced analytic capabilities for media organizations to employ. Readership demographics, preferences, and content trends are easily trackable through these platforms, which offer open web application programming interfaces (APIs), which provide windows to audiences.
The impact of the entrée of these new platforms is a dramatic shift in the way content and information is collected, produced and delivered. Content is now selected and shared among consumers (see Figure 2). In the process, consumers leave “digital exhaust” – data on the articles they select, how much of it they read, how long and how many times they re-read it, their location, and even what type of device they used to access the article. Their choices and preferences are instantly communicated to publishers, editors, journalists, and other content providers to help assess their responses.

THE RISE OF THE NEW MEDIA PROFESSIONAL

Powerful forces are converging and changing the way people consume information around them. More people are doing more things (interacting) with digital technologies, and increasingly, technologies are equipped to keep track of these interactions as data, representing human behavior and opinion. This has given rise to what is often referred to as “big data” – the availability of terabytes, petabytes, and exabytes of information in all forms – is enabling news consumers and professionals to dig into trends occurring across the globe, as well is their own backyards – to gain deeper insights and even make predictions on the future course of developments. The world is increasingly mesmerized by the growing array of social networking services available from the web. Global online communities with hundreds of millions of participants have emerged through sites such as Facebook, Twitter and LinkedIn. The “traditional” means of gathering news information – print media, broadcast radio and television – is rapidly being subsumed and disintermediated by new channels enabled by the web. No longer is information filtered by a few gatekeepers within centralized media sources.
News consumers are more empowered and have greater control, than ever before in selecting news services, contributing to the news, and promulgating news stories.

The enablers of this significant shift, digital technologies that support human interaction, the Internet, web, and mobile connectivity, have become primary vehicles in shaping today news services. The choices people make during interactions are instantly available to cadres of entities (journalists, editors, content providers, researchers, etc.) that can evaluate them and be responsive to the consumer needs.

This also opens up new kinds of roles for journalists, editorial professionals, and publishers. Today’s media professional is more than a purveyor of information and gatekeeper of information, he or she also needs to function partially as a data scientist, partially as curators of highly interactive and growing online communities, and full-time as critical thinkers who understand the value of questioning and revisiting the information they receive or uncover.

At the same time, there is greater demand than ever for the classic editorial skills that provide balance, vetting and critical thinking to content being delivered across digital channels. As anyone familiar with the Internet can attest, this network of networks is rife with inaccurate statements, rumors and unsubstantiated information. A paper published at the MIT Communications Forum quotes Thomas Cekay, who served as editor at The Chicago Tribune, who observed that traditional editorial roles – determining if readers

Figure 2. News consumer is more central in news services
need to see information, and if the information is intelligently presented – will increase in importance in the digital age. Information that is offered up by major media outlets still needs to meet certain criteria: its magnitude; its unexpectedness (which always makes it interesting); sociocultural values; continuity from previous content; and cultural proximity or relevance (Harper, 1998).

To a large extent, data-driven news and content still abide by these thresholds of what constitutes material of interest to audiences. The availability of data across many venues and activities opens up vistas for today’s media professionals. Data itself provides the raw material for news and informational reporting. Helping audiences to make sense of the petabytes worth of data now flowing through the public domain requires a high level of “storytelling,” in which data is analyzed and interpreted in ways that connect with audiences (Harper, 1998). To some degree, this requirement is the same that is found with the emerging category of data scientist, who is also charged with a storytelling role based on analysis and conclusions drawn from data sets. A description provided on the IBM website captures the nature of the data scientist role (IBM, 2014b, para. 3):

*The data scientist role has been described as “part analyst, part artist. Anjul Bhambhri, vice president of big data products at IBM, says, “A data scientist is somebody who is inquisitive, who can stare at data and spot trends. It’s almost like a Renaissance individual who really wants to learn and bring change to an organization.”*

Media professionals fit the mold as “Renaissance” individuals – capable of seeing the greater trends, of seeing incongruities or inadequacies in various situations, of being able to connect to their audiences. This data scientist-like storytelling role not only serves in being able to deliver data-driven content to audiences, but also for internal purposes – uncovering and evaluating data that provides insights on how deeply the content they publish is being absorbed by audiences.

The data increasingly are getting more detailed. At first, data from websites provided rudimentary feedback, such as the number of page views a particular article was receiving, and whether readers were responding to links to additional content, as well as advertising, on those pages. Now, tools and technologies are available to gain much deeper insights, reviewing audience members’ digital exhaust to provide real-time insights as to how readers are interacting with content.

It would be remiss not to mention the potential implications of automated article preparation, or what is affectionately referred to as “robot journalism” (Miller, 2015). The Associated Press is now employing algorithms that pull data from various press releases, including earnings reports, and formats the data into a readable news article. According to Miller, “the AP implemented the system six months ago and now publishes 3,000 such stories every quarter — and that number is poised to grow.” The AP reports were first closely monitored by human journalists before they were released to publication, but now, many are being released with minimal human intervention.

In an article in *New York*, Roose describes the typical process of generating an automated earnings report, and the various data feeds that make it possible. In this case, AP was employing software called Wordsmith from Automated Insights, to produce and publish the Alcoa earnings report:
When Alcoa’s earnings report hit the wire that day, the data was instantly compiled by a firm called Zacks Investment Research and passed through the AP’s proprietary algorithm, which pulled out key numbers and phrases and matched them against other contextual information. In milliseconds, the software produced a full story written in AP style, indistinguishable from one made with human hands. In short order, the story appeared on many AP member sites, including NPR, Hawaii News Now, and KAIT8. (Roose, 2014, para. 3)

Robot journalism has already been employed for the last few years in another data-intensive area of reporting, and that is sports reporting (Schonfeld, 2010). Such articles are produced in a matter of seconds, at minimal cost to the publishing organization. There are a range of potential publishing applications in which automated journalism will play an increasing role, including corporate financial statements, sales reports, and scientific findings. Automated report-writing algorithms may have a place in situations where audience sizes are too small to justify paying for the services of a human author. As explained by Kristian Hammond, co-founder and chief technology officer at Narrative Science, which developed authoring technology, in an interview with Marketwatch: “The place where it’s most powerful is where we have an audience of one,” he said. “Those stories tend not to be written by people, but they can be written by a machine” (Bemis, 2012).

While at first blush the specter of automated journalism appears to be a threat to the job security of journalists and editorial professionals, the technology may actually enhance and enrich the profession as it develops. Automated journalism has its limits, as it cannot interpret data and information within varying contexts, apply critical thinking, or be able to reach out to newsmakers or subject matter experts capable of providing perspective on what is being presented to audiences. Notably as well, algorithms cannot interact in a thoughtful and engaging way with consumers on social networks.

As Roose put it:

... these robots, as sophisticated as they are, can’t approach the full creativity of a human writer. ... My colleagues’ jobs (and mine, knock wood) are too complex for today’s artificial intelligence to handle; they require human skills like picking up the phone, piecing together data points from multiple sources, and drawing original, evidence-based conclusions. The stories that today’s robots can write are, frankly, the kinds of stories that humans hate writing anyway. (Roose, 2014, para. 8)

It will still be up to humans – journalists, editorial professionals and publishing executives – to apply the critical thinking that is essential to develop meaningful and engaging content for their audiences. These professionals still need to function as the “window” to events and information that may be inaccessible to audience members. Journalists and editorial professionals still need to connect and channel the thoughts of newsmakers and subject matter experts, while asking the hard questions that need to be asked.

CONCLUSION

Network connectivity, increased and ongoing human interaction with digital interfaces (devices), and the technical capacity to chronicle those interactions have been instrumental factors in the production of large amounts of data (big data) that news and publishing organizations can use to better understand
consumer behavior and ultimately enhance services and the user experience. Increasingly, more sophisticated analysis, tied to tracking technologies and analytic tools, give news organizations an even more in-depth view of their audiences. Through these new channels, consumers are also communicating directly, in both verbal and non-verbal ways, back to publishers, editors, journalists and other content creators.

William J. Gibbs  
Duquesne University, USA

Joseph E. McKendrick  
Independent Consultant, USA

REFERENCES


**Introduction**


KEY TERMS AND DEFINITIONS

Analytics: The analyses of both qualitative and quantitative data gather from sources such as server logs, data and text mining for the purpose of understanding trends and enhancing customer experiences.

Automated Journalism: The employment of algorithms to translate data and information into a human-readable article format.

Big Data: Large amounts of observational data, of different types, that are collected as user interact with computing devices. Often these data are collected in real-time.

Content-Sharing Platforms: Popular computing environments that allow people to exchange information in various media forms (text, video and audio, images) and to interact with other people worldwide in real-time.

Data Streams: Data that are collected in real-time. Technologies today enable researchers to record users’ interaction/behavior as they are happening.

Interface: A medium through which a human interacts with a system or machine.

Internet of Things: Refers to the connectedness of devices, platforms, and services from which vast stores of data from the Web, user-generated content, and social media can be obtained.

User Experience (UX): A user experience occurs when a person interacts with a system. It is influenced by internal characteristics of the user, attributes of the system’s design such as usability and functionality, and the context of use.