It’s not all cat videos: moving beyond legacy media and tackling the challenges of mapping news values on digital native websites

CANTER, Lily <http://orcid.org/0000-0001-5708-2420>

Available from Sheffield Hallam University Research Archive (SHURA) at:
http://shura.shu.ac.uk/23034/

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

CANTER, Lily (2018). It’s not all cat videos: moving beyond legacy media and tackling the challenges of mapping news values on digital native websites. Digital Journalism, 6 (8), 1101-1112.

Copyright and re-use policy

See http://shura.shu.ac.uk/information.html
IT'S NOT ALL CAT VIDEOS

Moving beyond legacy media and tackling the challenges of mapping news values on digital native websites

Lily Canter

The question "what is news?" has been a topic of scholarly examination for more than 60 years as researchers have sought to develop and revise the taxonomy of news values to inform journalism education and wider public debates on the subjectivity of journalists. However these studies have focused on legacy print and broadcast media even when attempting to bring their findings into the digital era by examining online content. To date there has been limited research, in the UK and internationally, on news values on digitally born news websites - platforms that only exist online such as Huffington Post, Buzzfeed News and LADbible which are consumed by a third of the UK population at least once a week. A significant impediment is the complexity of mapping offline latent coding methods to fluid online content. Digital native websites are particularly problematic due to their varied appearance across platforms - desktop, tablet and mobile - and their multiple access points - via web browser, social media and App. Yet content analysis has a rich history dating back to the clergy's examination of newspapers in the late 1600s and the method has proved to be an adaptable tool for measuring news output as each new media technology has emerged. This paper argues that it is imperative that researchers look beyond legacy media when studying digital news values due to the growing significant of digital native news websites in the marketplace. The secondary purpose of this methodological paper is to highlight the challenges of capturing and analysing news values on digital native news platforms and suggest how researchers can begin to tackle the complexities of liquid content analysis in this field.

KEYWORDS Content analysis; news values; digital native; research methods

Introduction

The study of news values, the criteria by which journalists select events or information to become news stories, has a century of rich scholarship spanning the globe. In 1922 political commentator Walter Lipmann argued in his influential tome *Public Opinion* that the public was not fit to make decisions about newsworthiness and it was the job of the newsman to shape the public's view of the world (Lipmann
His thinking was further developed in the 1950s when the topic of news selection became increasingly important in the field of communication studies. Six news factors began to emerge from American journalism textbooks which defined news as being related to directness, proximity, prominence, unexpectedness, conflict and significance (Staab 1990). But it was not until the 1960s that news selection factors began to be analysed empirically with results showing that the more news factors a story contained the more newsworthy journalists considered it to be. European research established by Ostgaard (1965) developed a complex theory that condensed different components of news into three factors: simplification, identification and sensationalism. His theory was developed by Galtung and Ruge (1965) whose seminal paper remains one of the most cited to date despite its shortcomings. Galtung and Ruge used a systematic content analysis to explore the international reporting of conflict in the Congo, Cyprus and Cuba and devised 12 news values which formed the basis of further empirical studies throughout the 1970s and beyond (Staab 1990). The limitations of Galtung and Ruge’s study were most vehemently critiqued by Harcup and O’Neill in their equally significant 2001 paper which argued that “despite the way it has been so widely cited, Galtung and Ruge’s taxonomy of news factors appears to ignore the majority of news stories” (276). The authors conducted their own content analysis of 1276 news articles published in three national newspapers, testing and modifying Galtung and Ruge’s taxonomy to develop 10 contemporary news values. They subsequently updated this list in 2016 following a wider analysis of 10 newspapers together with a preliminary examination of stories with high sharing metrics on Facebook and Twitter (Harcup and O’Neill 2016).

Content analysis has formed the backbone of this type of empirical news values research with much of it being coded from print newspapers (Galtung and Ruge 1965; Schultz 1982; Harcup and O’Neill 2001; 2016). However this fertile field of research has also produced systematic studies examining television news (Bergsma 1978; Buckalew 1969; 1970), radio (Niblock and Machin 2007) and photographs (Craig 1994; Singleterary and Lamb 1984). Furthermore news values are a universal phenomenon explored by scholars on an international scale with studies conducted in Argentina, USA, UK, Sweden, Denmark, Germany, Australia, South Korea and beyond (Craig 1994; Lee and Choi 2009; Schultz 2007; Wendelin at al 2015). But the canon of research in the field is far from exhaustive as there still remains a significant gap in understanding around news values and online news, particularly digital native news websites. Instead researchers have continued their preoccupation with legacy media by analysing differences between print newspapers and their online output (Strömbäck, Karlsson and Hopmann 2012; Welbers et al. 2015; Wendelin, Engelmann and Neubarth 2015), comparing digital native websites with legacy media (Tandoc Jr 2017) or focusing on news shared on social media (Hoffman 2006; Bednarek 2016; Harcup and O’Neill 2016; Garcia-Perdomo et al. 2017) again predominantly coding the online content of traditional media organisations. And while Küng (2015) has strived to understand and compare the
corporate strategy and cultural processes within digital native news websites and legacy media, and Eldridge II (2017) has examined the journalistic identity of content creators on emerging platforms such as Wikileaks neither have explicitly analysed news values.

This article seeks to explore how content analysis can be used as an effective method for understanding news values on digital native websites. It attempts to put forward a strategy for overcoming the problems that arise from dealing with content known as liquid journalism (Karlsson and Stromback 2010) which is fluid in both time and space.

Rise of Digital Natives and Challenging Elitism

Although news values research to date has analysed a range of media platforms including online, these studies have concentrated on legacy media and all but ignored digital native news websites. These are news outlets that originated online and are not born out of legacy offline media. News outlets such as Huffington Post, Buzzfeed News, LADBible, Breitbart and The Canary now share a significant part of the online news market with a third of the news-consuming UK population accessing them at least once a week (Newman 2016). In Australia and Japan digital native news sites are the market leaders with two thirds of the population accessing them, while just a third consume newspapers and broadcast media (Newman 2016). The Huffington Post is currently the fourth most popular online news source in the UK, sitting just behind The Guardian and Mail Online, with Buzzfeed News placed at seventh and LADBible at eleventh with six per cent of the market share (Newman 2017). With their growing audiences and consolidation within the market these digital native news websites form an essential part of the news values ecosystem and yet they are rarely discussed in news value literature (Garcia-Perdomo et al 2017). Instead scholars continue to fixate on traditional media despite repeated studies suggesting that their online content simply reinforces their offline legacy platform (Hoffman 2006). A case in point is the revised news value taxonomy of Harcup and O'Neill (2016) which sought to gain a greater understanding of the shareability of stories on social media. Their analysis of most shared stories was based on a list of the ten most popular online news stories from mainstream media, extrapolated from journalism.co.uk and Newswhip. But digital native news websites were not included in this analysis despite being known for their proliferation on social media.

With the millennial generation consuming more news from digital native websites than any other age group (Newman 2016) is it vital that researchers look beyond legacy media to understand how news values are evolving. O'Neill (2012) berates the malaise of quality newspapers in her noteworthy research on celebrity news values and asks where will the sections of the public interested in serious news, comment and analysis get their content in the future if The Daily Telegraph,
The Times and The Guardian continue to embrace “the banal, gossipy, celebrity content of the tabloids” (2012, 28). But what the author fails to acknowledge is the existence of alternative news media and the vast array of digital native news organisations that are offering this quality content. Millennials are already turning to these digital native platforms rather than legacy newspapers (Newman 2016). As researchers we ignore these news providers at our peril as in order to understand the future of news values and the practices of journalists we must interrogate the newest players to the field.

One unsubstantiated justification for the lack of investigation into news values on digital native sites is the assumption or misconception that these websites contain soft news and are proliferated with “funny cat videos” (Bednarek 2016, 232). The vacuum of empirical data in this field appears to indicate an elitist attitude towards digital native news organisations which implies that their content is not worthy of academic investigation because they do not produce 'real' news. This is despite evidence that journalists working for traditional news organisations positively welcome Buzzfeed's entry into the journalistic field and see it as reinforcing existing professional norms (Tandoc Jr and Jenkins 2017; Tandoc Jr 2017). Therefore without systematic analysis there is no evidence that news on digital native websites is in some way not credible or unsubstantial. How do we know that that the content "cannot traditionally be called news" (Bednarek 2016, 232) if we do not hold these websites up to academic scrutiny? News values research of digital native news websites is a step towards this understanding. We may think we know the content of digital native news platforms but it is only with systematic analysis that we can actually begin to know.

The Problem with Liquid Content Analysis

Content analysis has featured heavily in news values research for centuries and was implemented as far back as the 1600s when newspapers were examined by the church in response to concerns over the spread of non-religious matters (Krippendorff 2004). As a systematic, objective and quantitative method it has proved to be a robust means of analysing complex, static data such as the content of newspaper pages and television news programmes and in more recent years has emerged as an effective tool to research the content of web pages and websites, including user comments (Trice 2011) and user generated content (Jonsson and Örnebring 2011). But much of this prior web research was developed during the eve of online news when content remained largely static “with a set beginning and end” (Karlsson and Sjovaag 2016, 181) similar to offline analogue media formats. Today the content is fluid, with news stories and news pages being updated at irregular and unpredictable intervals. An additional barrier is that these websites often contain multimedia content which cannot be captured consistently by open source resources
(Hurwitz et al. 2006) particularly Flash content or Graphics Interchange Formats (GIFs) which are popular on digital native websites (James 2014).

Due to the complexity of studying constantly changing new media content researchers should consider how to adapt existing content analysis tools (Harcup and O'Neill 2001; Strömbäck, Karlsson and Hopmann 2012; Tandoc Jr 2017; Welbers et al. 2015; Wendelin, Engelmann and Neubarth 2015), and assess what can, and cannot, be imported into the digital realm. This requires a customised approach which will differ with each digital news organisation under examination and requires alignment with the research goals of the particular study.

Here the author reviews and discusses some of the challenges associated with capturing and coding news values on digital native websites before offering an example of some of the ways in which these can be overcome in reference to a specific pilot study.

**Fluidity of content**

Conducting a content analysis of news values in a newspaper is a relatively simple process due to the static nature of the content and the entry point most commonly being the front page. It is also fairly easy to distinguish between editorial and advertisement. A different set of challenges appear when approaching a content analysis on a digital native website. Firstly researchers must decide what type of content is appropriate to capture. Should the data set include blogs, promotional stories, advertisements, advertorials, opinion pieces, features or multimedia posts and if not, is it possible to separate news stories from this other content?

The next set of issues relate to time and space which are acutely fluid in the online arena. How often does the content change and when is the most appropriate time for a researcher to capture it? Should it be captured at regular intervals to adequately observe an unfolding story? Should it be captured at peak traffic time or at a period when the content is more stable and less likely to change? How can it be captured quickly? Even more difficult to approach is the issue of where to capture content as it becomes "increasingly difficult to pin down where the news is produced, distributed and consumed as news appears in so many places and comes in so many forms at once" (Karlsson and Sjovaag 2016, 180). Digital native news websites are accessed via web browsers or Apps on desktop computers, mobiles and tablets as well as via social media platforms such as Facebook and Twitter. Users can enter via the homepage but hyperlinks on social media platforms or search engines will connect users directly to a story's individual web page which does not necessarily feature on the digital native website homepage. A coder therefore needs to decide whether it is relevant to capture stories on the homepage or on other content pages (for example News or Politics) and how many stories to capture because the page may scroll down indefinitely. Websites with sophisticated algorithms like *Buzzfeed News* will have different stories on the homepage of each of their platforms (desktop,
mobile, tablet) and there will also be disparity between the App and the web browser and between Android and iOs devices leading to multiple variables to consider.

Sample size

Once a capture point has been identified in time and space the amount of data to be sampled must be decided. How much is enough and how regularly should it be sampled? If the coding context is a homepage on a mobile browser how many of the stories should be captured and coded? And should they be captured more than once a day? Another consideration is whether to sample consecutive weeks or constructed weeks. A review of online news research using content analysis shows limited evidence of the application of random sampling as an effective way of gaining representativeness and generalisability (Hester and Dougall 2007). They advocate for a minimum of two constructed weeks but reason that as many as five may be needed. Furthermore Hester and Dougall demonstrate that consecutive day sampling performs poorly and "should probably be avoided" (2007, 820).

Overcoming the challenges of liquid content analysis

The problems outlined above although taxing, are not insurmountable. A researcher must make a set of decisions to answer a complex set of challenges and in lieu of a foolproof solution choose the next best thing and stick with it. This means "embracing the messiness, promoting design transparency and supporting kludginess" (Karpf 2012, 642). This article now aims to explain how the author overcame the challenges of measuring news values on the three most popular UK digital native websites: Huffington Post, Buzzfeed and LADbible, as a lone researcher with no budget.

Defining coding unit

Testing and developing an existing taxonomy of news values can be an effective way to use a combined deductive - inductive approach. In order to analyse news values on digital native websites the author devised a coding unit which categorised the 15 news values set out by Harcup and O'Neill in 2016 plus an Other category to allow for a more inductive approach and the potential for an evolving taxonomy. This clearly places the research within the cognitive perspective of news values viewing them as a set of judgements made by journalists (Bell 1991; Kepplinger and Ehming 2006; Stomback et al. 2012). This interpretation rationalises news values as characteristics of journalists rather than as qualities of news stories. If we wish to understand what decisions journalists make and which stories they deem most significant then we need to capture the content that takes priority on digital native news websites via the order it is ranked in. Coding the most viewed or most shared stories will not tell us about the judgements of journalists but only shed light on the judgements of consumers. If news values are intrinsic to understanding
the practices of journalists then even in the realm of liquid journalism we must interrogate the original story.

**Defining coding context and capturing fluid content**

Pilot studies and exploratory testing are vital to ensure a robust coding context and capturing method are secured in such a fluid environment. The home page or main news page of a news organisation website can be an ideal unit of analysis as this is where journalists will set out their top stories in priority order thus displaying the news values they judge most valuable. This will be complemented by the social media feeds of the organisation which may promote a different set of stories and therefore via a pilot it must be determined whether there is value in simultaneously analysing this content or whether it closely matches the stories on the website home page. With this in mind it is imperative that the coding context is aligned with the research goals and framework of each individual study. It may be more important to determine the range and nature of content that ultimately reaches audiences and therefore coding most viewed or shared content may be more appropriate or indeed the content selected by journalists to distribute on social media may also be worthy study.

Once an appropriate coding context has been chosen then a researcher must decide whether to include non-news items and should carefully research the websites selected to understand their content types. The author would advocate the dismissal of any material signposted as promotional, advertisement or blog. For example on the Huffington Post website promoted content is marked with a banner which includes the name of the sponsor (see Figure 1) so this content would be removed from a data set before coding.

**Figure 1: Huffington Post Promoted Content**

![Huffington Post Promoted Content](image-url)
Liquid content analysis is often seen as a means to capture constantly updating content and to tackle the “flowing river” of online news (Karlsson and Sjovaag 2016, 186). But the assertion that immediacy is a fundamental trait of online news has been demonstrated as a mythology as institutions have established routines and rules for updating content (Lim 2012) and there is only a low level of immediacy. Saltzis (2012) demonstrated that news stories do not change much after the initial hours of publication. Furthermore, a pilot conducted on the digital native websites in this study also found that during a four-hour, weekday period from 8am to 12pm only one story was updated across the three sites and this level of immediacy was similar in subsequent trials.

However, a more significant problem which has received little attention from scholars is the disparity between different platforms and access points of the same news provider. A pilot for this paper demonstrated that the Buzzfeed mobile App and tablet App did have the same content, in the same order as the mobile web browser, but this content differed to the desktop and tablet web browsers (see Figures 2 and 3).

Figure 2: Top three stories on *Buzzfeed News* desktop web browser

Figure 3: Top three stories on *Buzzfeed News* Android mobile web browser
Furthermore research tells us that access to news via mobile is increasing and via desktop is decreasing (Newman 2016). Millennials, those aged 20 to 35, who are the target audience of digital natives, access news via mobiles at a significantly higher rate than tablets or computers. It is therefore rational to suggest that content for digital news value analysis should be captured from a mobile browser as this is likely to be the most popular entry point for users (albeit via a side entry point such as social media) and thus the focal point for journalists, with Apps seeing little growth in recent years (Newman 2017). It is also easier to navigate the mobile web browser as a researcher because the content is more clearly signposted and presented in a neat, vertical list rather than across unaligned columns. For example on the Huffington Post mobile web browser the content is clearly marked as 'Blog' but on
the desktop web browser this content is mixed amongst news content and not clearly indicated as non-news.

**Sampling news items**

With the context of the time and place identified the final step is a decision on how much content to capture and for how long. A pilot of the three sites in question indicated that a sample size of the top 20 news stories from each News home page was an appropriate number. For example this is the total amount of stories displayed on LADbible before users have to press the 'More' button and it is also a large enough sample that at least 50 per cent of the content is the same between BuzzFeed desktop/tablet web browser and the App/mobile web browser. It is worth noting that on BuzzFeed a sample size of 40 would be required to obtain a 100 per cent match rate between platforms/access points but users are unlikely to scroll down the page this far and due to the rationale above it was deemed that the App/mobile browser was more relevant as it targets the key millennial audience. If 20 stories were captured on each website once a day during three random constructed weeks across a six month period this would generate 1260 stories for coding. This is a similar amount to the 1276 articles analysed by Harcup and O'Neill in their original 2001 study.

**Conclusion**

In the realm of news values research it is time for scholars to step away from the comfort zone of traditional news organisations and look towards digital native news providers to look objectively beyond the "funny cat videos" (Bednarek 2016, 232). The mainstream media is receiving mounting criticism over its failure to connect with the electorate (Harrison 2017) and as such users are turning to "alternative voices" for their news and comment. It is therefore imperative that researchers do not fall into the same trap by becoming preoccupied with mainstream media and ignoring the growing number of alternatives.

The author advocates an inclusive approach towards online news values research and the incorporation of digital natives into the evolving narrative. There is a desperate need for more empirical data on non-traditional news outlets which are rapidly becoming the news providers of choice for the millennial generation. As researchers we must be willing to embrace these new forms of communication and subject them to the same rigorous testing as we do legacy media.

In order to do this researchers need to adopt a malleable approach to content analysis to transfer latent coding methods across to the fluid landscape of online liquid journalism. The key to a successful research design is a thorough and systematic understanding of the news organisations under interrogation and the implementation of a transparent research process which has been developed.
through trial and error. Pilot studies are a key component of this design as they enable researchers to explore the uncharted territory of news values on digitally born news providers and develop a method to make sense of the multiple platforms and devices on which they publish news in response to the specific research goals. The liquidity of these websites is connected to much more than immediacy and multimedia content and a greater challenge is how to tackle their illusive bespoke algorithms and the multiple access points and technologies through which their content can be consumed.

This paper is the starting point for further discussion on how to devise effective tools for measuring news values on digital native websites and offers some solutions whilst also raising many more questions. It is the intention that this will be the jumping off point for further research, perhaps with the help of computational studies colleagues, to truly understand the content of these popular, emerging news platforms.

References

Bednarek, Monika. 2016. "Investigating evaluation and news values in news items that are shared through social media". *Corpora* 11 (2): 227-257.


Hoffman, Lindsay. 2006. "Is internet content different after all? A content analysis of mobilizing information in online and print newspapers". *Journalism and Mass Communication Quarterly* 83 (1): 58-76.


