Coding Twitter, lessons from a content analysis of informal science

Noah Goodman (ngoodman@edc.org)

Daniel Light (dlight@edc.org)

March 18, 2016

EDC | Center for Children and Technology

Twitter has become an important platform for textual communication and information sharing. With 500 million tweets sent daily, it offers an abundance of accessible and economical data on human interactions. But tweets exhibit specific characteristics of brevity, fluidity, and meaning embedded in a broader context, which pose serious challenges for the researcher engaged in content analyses.

In this paper, we would like to share our experience conducting a content analysis of Twitter data as part of the TwISLE project, a three-year study funded by the National Science Foundation to investigate informal science learning and engagement on Twitter. We will use de Beaugrande and Dressler's standards of textuality (1981) to highlight peculiarities of tweets as a form of text, explore the challenges these raised for our research process, and suggest solutions for rebuilding the context that is often elusive on Twitter.

Figure 1: #cephalopod

PERSPECTIVE

Our project was born from an interest in the ways people encounter science on Twitter. We sought to investigate what types of information were being shared, what interactions were being represented, and how this platform related to a larger infrastructure of informal science learning. We also wondered about the roles played on Twitter by institutions such as natural history museums, aquariums, and zoos. As the public-facing representatives of science, they seemed well suited to a social platform like Twitter.

As we spent our mornings reading streams of tweets, we found ourselves huddling with friends around a video of a #cephalopod camouflaging itself on the sea floor, telling our partners how curators used 3D printers to create bones for the new #titanosaur on display @AMNH, and pondering the ingenuity behind @MontereyAq's integration into the natural world as we watched a wild otter choose the aquarium's tide pool in which to birth her pup.

Figure 2 - #Titanosaur

We sensed layers were being added to our relationship with the world of science and the institutions that represented it. Each tweet became a tiny building block, its contribution leaning on that of the previous one. It was the totality of these minor interactions with information and individuals that came to define our experience on Twitter.

Over the years, Twitter users have taken a fairly simple platform and innovated ways to engage each other and topics of interest. The two primary features that are used to organize

communication are the @username and #hashtag conventions.

While both of these conventions are also used in discursive ways, their primary significance is as searchable text (Zappavigna, 2011). The @username is the basis for "sender-audience" relationships making "communicative references to other Twitter users not only visible, but navigable as well" (Schmidt, 2013, p. 5). In the tweet A short piece on sea otter tool use in California, featuring Jessica Fujii and her team at @MontereyAq https://t.co/5jNSqJq9Nt, @MontereyAq works a double service. It alerts the Monterey Bay Aquarium that someone has @mentioned them in a Tweet. It is also clickable, allowing a user to navigate to the Monterey Bay Aquarium's profile.

**Figure 3: Conversational thread - the rescue of a sea lion**

The @reply is a specific type of mention (Twitter, 2016) that is the basis of what become conversational threads such as the exchange in figure 3 around the rescue of an injured sea lion.

Users receive updates from others by following them, and their timeline features a reverse-chronological stream of these tweets. But they can also access information by searching @mentions or #hashtag conversations.

Hashtags are keywords, like #citizenscience or #paleontology, with which users tag their tweets. A tweet with the hashtag #citizenscience enters into a topical stream accessible irrespective of follower/followee relationships.

In the interviews we conducted with followers of these public science institutions, users reported gaining access to specialized science conversations on Twitter. Following a paleontologist or a cosmologist might open whole domains of information to the curious user. However, finding these scientists on Twitter requires a level of awareness of the scientific community that the public often lacks. More visible accounts—celebrity figures like @BillNye and @neiltyson or institutions like @FieldMuseum or @Bronxzoo—can be points of entry. Reading tweets from these accounts might lead the user to find new hashtags such as #trilobite and #entomology or figures such as @CorrieMoreau (evolutionary biologist @FieldMuseum) or @CoralReefFish (curator of fishes @CalAcademy).

Users themselves can also become gatekeepers of scientific information regardless of their professional background. Several people we interviewed spoke of the audiences they kept in mind when they tweeted scientific information. They sent out science-related tweets they felt would "benefit people," that would "dispell myths" their followers might have, or that would lead them into having "a more interesting water cooler conversation." And they were excited by the idea of sharing this with followers who might "not have been inclined to go look for information."
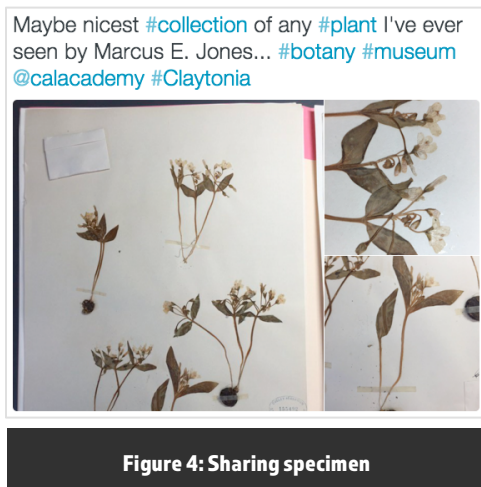
Figure 4: Sharing specimen

As people interact with the individuals and organizations invested in a topic, they build in their Twitter profile a public-facing display of their expertise. Similar to children guiding their parents through a museum exhibit about which they are knowledgeable, this type of information sharing "builds up a sense of self as one who is knowledgeable" (Barron, 2006, p. 198). And as others begin to look to them for this type of information, that identity is

externally reinforced. St. John and Perry (1993) suggest that museums can be a place where people "meet" science. We suggest that Twitter can be a place where people "friend" science, and where they can have an OMG @nasa replied to me moment.

Our content analysis sought to capture not only the information that is shared through these interactions, but also the serendipitous and unexpected aspects of the Twitter experience. As we gathered our data into spreadsheets during our initial coding attempts though, we quickly realized the text in front of us felt detached from these hard-to-quantify, affective elements.

In the following sections, we will explore aspects of this detachment by viewing tweets through the lens of three standards of textuality.

OUR PROJECT

The primary goal of this project was to develop two research instruments that could be used to investigate informal science learning on Twitter: an informal science engagement (ISE) index based on a content analysis of the conversations surrounding public science institutions on Twitter, and a qualitative survey that could be used to collect data from the followers of these institutions about their motivations and their experience of informal science on Twitter.

We started collecting data from a range of public science institutions, including government agencies (@NASA and @NOAAClimate), museums (@AMNH and @FieldMuseum), and zoos (@BronxZoo and @ClemetZoo). As the project progressed, we expanded the focus to more place-based institutions—realizing that their streams were significantly different from those of the government agencies—and we added @MontereyAq and @CalaAcademy to our radar.

Since were interested in investigating the nature of the conversation surrounding these institutions, we wanted to look at not just official outgoing tweets but also @mentions of the organizations. In order to do this, we built a dataset for each organization around the search query @username OR from:@username (i.e., @AMNH OR from:@AMNH).

TWITTER AND TEXTUALITY

Social networks such as Twitter, Facebook, Instagram, and Tumblr are sites of networked publics. They are a locus of social interaction within a specific technological architecture, an architecture that includes certain affordances that "do not determine social practice," but that "can shape publics and how people negotiate them" (boyd, 2010, p. 7). The product of this interplay between social action and technological affordances has produced a type of text—the tweet—with specific characteristics of brevity, fluidity, and meaning embedded in a broader context.

Bruns and Moe (2013) have suggested that communication in Twitter exists simultaneously on three levels: macro, which is completely public—even available to individuals without Twitter accounts; meso, in which communication through follower–followee networks is part of a group conversation; and micro, in which a tweet is directed to a specific individual or to institutions, such as .@AMNH what time do you open today? Exchanges usually take place across several levels at once, and the addition of a hashtag or subtraction of an @username can easily shift the level of conversation. As the sender modifies the audience for their tweet—or negotiates multiple audiences—the relevance of contextual markers shifts.

Conducting a content analysis was our attempt to look across these three levels and use the traces of human communication to recognize salient elements of the user's experience of informal science on Twitter. Our coding scheme became our API for connecting to this experience: a set of explicit guidelines and definitions that enables us to

assess the tweet in relation to the larger conversation. We came to see every tweet as a tiny building block within a larger conversation, but whose meaning was interconnected in a way that was often not be visible to us.

De Beaugrande and Dressler have proposed seven standards of textuality that ensure successful communication. Weakness in any of these elements will leave the reader struggling to apply proper meaning. But these elements were developed for more traditional texts, such as newspaper articles, journals, and letters, which are often physically bound in their context. In this section, we use three of the standards—cohesion, situationality, and intertextuality—to explore peculiarities of the tweet as a form of text, and some of the challenges we encountered working with tweets as units of analysis.

Cohesion. Cohesion concerns itself with the grammatical and lexical elements that connect thoughts into comprehensible messages. Twitter's 140-character limit imposes serious constraints on the messages users share. In order to economize length while maximizing meaning, users often employ abbreviated or unconventional spellings (Han & Baldwin, 2011). The inconsistent grammatical structures that emerge pose difficulties for researchers attempting to infer meaning or create coding categories that rely on the presence of certain grammatical elements, such as an active verb.

Some tweets we have gathered, like @NASA What more my antenna optic have to vein judicial @X ovary century? seem fragmentary or nonsensical, and we have deemed them uncodable. Other tweets, like M1.2 - 28km NNW of Packwood, Washington http://earthquake.usgs.gov/earth… 1.2 12th Event, contain relevant content but present themselves as cryptic puzzles requiring the researcher to rely more heavily on contextual elements, such as the embedded link—which in this case clearly shows that the tweet is in reference to an earthquake.

Situationality. The meaning of a text is often tied to the time and space in which it's presented. However, as is noted by Bruns and Moe, the levels of communication on Twitter

are fluid, and any individual tweet will appear to a user randomly sandwiched between millions of others. Context in Twitter is an elusive element when compared to other forms of computer mediated communication, such as forums, comment threads on blogs, or Facebook walls, where a post will more likely stay connected to its referents. On Twitter, the situation of the reader is easily disconnected from the situation of the sender.

The case of communication at the micro level through @reply conversations presents ample examples of the effects of situationality. Bruns and Moe use the analogy of an offline conversation between two friends but in the presence of a group of people to describe these exchanges. But what if someone listening in that group were to go home and tell her partner about the conversation, who then went to work and told his colleagues about the conversation. Even if the initial message remains intact, each new situation has the potential to influence the way the it is understood.
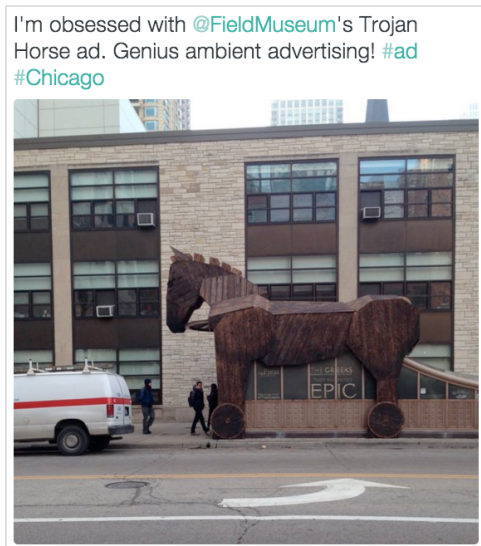

I'm obsessed with @FieldMuseum's Trojan Horse ad. Genius ambient advertising! #ad #Chicago

Figure 5: #Situationality

At the meso level, situationality can affect meaning as #hashtags are added to tweets, thus inserting them into conversations that are often as diverse as they are similar. The tweet *#3Dprinting used by @NASA to construct first copper rocket engine part http://t.co/KEBEe6dTik* sends the message into a conversation about 3D printing that might also include tweets about fashion or design. That tweet simultaneously lives in a stream of @NASA mentions and is probably understood differently by users following these different spaces.

The practice of retweeting is another process by which tweets are often removed from their original context and brought

into new ones. In the literature on Twitter, this is both celebrated as facilitating "serendipitous discovery of interesting ideas and people" (Halavais 2013) or condemned as disconnecting academic conversations from the original texts they referred to (Gallezot 2013). Gallezot is concerned that critical academic conversations about new scientific discoveries on Twitter easily become disconnected from the actual article being debated—leading to confusion and misunderstanding.

The fluid situationality in Twitter means that it can be a struggle for researchers—like Twitter users themselves—to apply the appropriate meaning. The appearance of these tweets in a dataset can create anomalies that research teams must decide how to deal with.

Intertextuality. Intertextuality highlights how a text takes meaning from its antecedents. If the receiver of a text is unaware of those proceeding it, her or his ability to establish meaning will be inhibited. On Twitter, the emphasis is equally on the surrounding content—including text as well as video, images, and even offline events and conversations.

Recognizing and applying thematic content analysis codes to tweets without these surrounding texts is like being in a loud party where everybody speaks a language in which you're merely proficient. Although we can accurately code around two-thirds of our data without referencing anything but a tweet's text, it is very difficult to assign specific content codes to that final third without access to externally referenced content.

We have identified five forms of intertextuality on Twitter that affect researchers' work: tweets have embedded media such as images, video, or audio; they use links to point to external websites, articles, or media; they refer to other tweets within a conversational thread; they refer to temporal events; and they are part of specialized conversations. In each of these instances, meaning from prior and surrounding texts is often implicitly incorporated into the tweet being coded.

Whereas the text of a tweet often relies on embedded media to convey meaning, the process of analyzing tweets as data generally involves discarding that media. In the figure below, for example, it would be hard to understand *@nasa #girlswithtoys* without the accompanying image of the famed female pilot. While the sparse text of this tweet exemplifies how meaning is built across multiple elements, the hashtag #girlswithtoys—generally applied to tweets featuring women in the STEM fields—is itself a nod to layers of intertextuality.

Just as the previous tweet illustrates how embedded media can be used to create meaning, the tweet #JarJarBinks took a job at the @FieldMuseum after the #Episode1 fiasco. #StarWars #Chicago [URL] emphasizes how the absence of embedded media affects our ability to reliably apply thematic codes.

Figure 6: #intertextuality and embedded media

We are interested not only in how Twitter constitutes a unique arena in which to experience science but also in how this arena connects to other spaces. Our interaction code thus captures representations of users' museum visits. Embedded images, such as the one accompanying the #JarJarBinks tweet of a person standing in front of a dinosaur exhibit, are often the only distinguishing factor between a tweet that mentions a museum and a tweet about a visit to a museum.
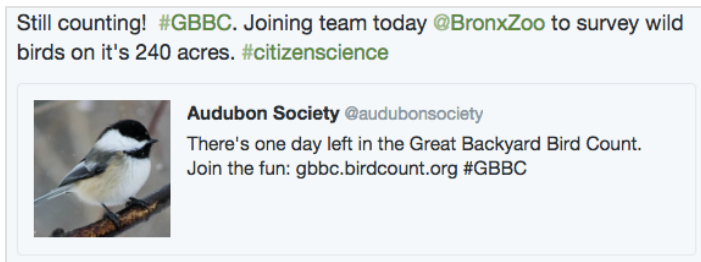

Figure 7: Retweet with a comment

We use the term embedded media to refer to images, videos, and audio files that appear alongside the original posting of a tweet. This

is information that a user on Twitter would see without having to click on an external link. As a result of two newer features on Twitter, however, text can also be embedded so that it is visible to the user but hidden from a researcher's dataset.

In April 2015, Twitter added the "retweet with a comment" feature, allowing users to add a comment of up to 116 characters to their retweet (Shu, 2015). The resulting tweet appears to users with the comment on top and the retweet—including embedded images—below. In a dataset retrieved through the API, on the other hand, only the text from the comment appears, followed by a link to the original tweet.

The other feature that allows text to be embedded without appearing in an archive is Twitter Cards, added in 2012. Twitter Cards allow content producers to use meta tags on their website to designate specific media and summary text to accompany tweets of their content (Roomann-Kurrik, 2012). In these instances, a dataset may show the presence of a URL but no indication that embedded media or text accompany the tweet. Furthermore, since the information on Twitter Cards is designated at the site of the external link (Twitter, 2016), users themselves do not know this information will be included until they've sent the tweet.

Figure 8: @reply

A majority of tweets we have captured in our dataset feature links to external websites. These links represent another form of intertextuality. Some tweets with links, such as @NOAAClimate: Rainfall across parts of #Texas and #Oklahoma was WAY above normal in May 2015. Our update: http://t.co/Lf8w5XLWIm

http://… offer clues to the contents of the link. Others rely more heavily on the content of the link to construct a coherent unit of meaning.

The tweet Twilight in Wired: Great to have been part of this in 2013-14 @calacademy @deepreef @the_deep_greene @CoralReefFish : http://t.co/mwcRRZhGOb links to a story about scientists from the California Academy of Sciences studying deep sea coral reefs. We are interested in any mention of scientists or the work they do, but following the link was the only reliable way we could have applied our Being a Scientist code.

Other tweets include either nothing but a link or a link with a series of hashtags, such as @montereyaq #montereybayaquarium #monterey #california #CA [URL]. The tweet links to a picture on Instagram that the user took, presumably during a visit, of a fish on display at Monterey Bay Aquarium.

Another form of intertextuality is made up of tweets that refer to other tweets through @replies or conversational threads. Reply tweets often carry implicit meaning from previous turns in the thread. Beyond the challenge of constructing meaning, this can create real questions about how to code what is carried over from the original tweet. If @AMNH tweets Get ready, this feathered #FossilFriday is part a new Museum exhibition, opening next week! http://trib.al/LAOu9eG—a tweet that we would code as mentioning an exhibit—should the @reply @AMNH will it be scanned at some point and made available for 3D printing at home or school? also be coded as mentioning an exhibit?

Conversational threads are generally fairly short, consisting of one or two @replies (such as in figure 8), but they can become lengthy, especially as new users are brought into the conversation through the use of @mentions. The sea lion conversation (figure 3) is an excerpt from a thread that spanned 42 tweets from nine individuals and two institutions over the course of four days. As conversations lengthen and the subject evolves, the task of attributing implied meaning becomes more and more complicated.

Twitter was originally designed for use on mobile phones—hence the 140 character limit, corresponding to the text allowed in a SMS message. Although desktop apps are now widely used, Twitter is still a platform that is regularly integrated into people's lives as they move from one situation or location to another. Users tweet not only about events but also from events, representing a fourth form of intertextuality.

The tweet Everyone should eat more bugs! From one #amnhscicafe to another. @AMNH seemed like a random joke when we first saw it. After further investigation—by searching the hashtag #amnhscicafe on Google, viewing the profile of the sender, and finding the original post on Twitter's website—we learned it was about insects as a food source, was sent by scientist working with the American Museum of Natural History, and was tweeted during a lecture he was giving at the museum during an event called SciCafe.

Finally, Twitter is the site of many specific conversations with their own vocabulary, memes, and hashtags that are not always accessible to the outside reader. Fluid channels of dissemination on Twitter mean that these conversations—which on other platforms might be more consistently siloed in specialized forums or groups—easily enter the broader public sphere.

The tweet #LookUp #VA #DC #MD 10/24 as @EPA @NASA @NOAA LINEourSKYwith #Chemtrails #Geoengineering #SolarRadiationManagement http://t.co/Kg6jUpsIje is unintelligible to a reader unfamiliar with its references, but it fits into common memes in the @NOAAClimate Twitter stream from users skeptical of climate science. #Chemtrails, #Geoengineering, and #SolarRadiationManagement refer to a belief in secret government run geoengineering programs; #LookUp suggests that the answer is in the sky; and embedded images included in the original tweet show thin trails of clouds running across the sky—the supposed product of seeding chemicals into the atmosphere.
This tweet is also a good example of how situationality can affect meaning. Since its author has included the hashtags #VA, #DC, #MD and mentioned @EPA, @NASA, and

@NOAA, any users or researchers following information about any of these states or organizations might have this tweet appear in their feeds.

The intended meaning of a tweet is often wrapped into a complex exchange between technology, human interaction, and the culture that has evolved around Twitter. Since content analysis has become the dominant method used for studying Twitter data (Zimmer & Proferes, 2014), it is important to develop strategies to peel back these layers during the coding process. In the next section, we will discuss steps we took to address some of the challenges posed above.

REBUILDING THE CONTEXT

As we have seen, coding for meaning with nothing more than a tweet's text often leaves a researcher in the dark and disconnected from the experience of the user. In our coding process, we have used several pieces of metadata that allow us to rebuild layers of contextuality.

A given tweet may contain up to 150 pieces of metadata (Dwoskin, 2014), including everything from how many people follow the tweet's author to the program used to send the tweet. Researchers can capture tweets and their metadata through a number of applications that connect to Twitter's API—either in real-time using the Stream API or in the past through the REST API.

Numerous researchers have written about using the now-defunct TwapperKeeper (Zimmer & Proferes, 2014; Williams, Terras & Warwick, 2013; Gleason, 2013), and Bruns and Liang (2012) have discussed using an offshoot program called yourTwapperKeeper. After trying several paid and free applications, we happily landed on TAGS V6.0, a free and open-source project that utilizes a customized Google spreadsheet to connect to either of these APIs.

By default, the TAGS spreadsheet includes 16 columns of metadata. Of these, we primarily use two: the in_reply_to_status_id_str and the status_url.

from	text		in_reply_to_status_i	status_url
bion	@MontereyAq Wow! I'm assuming that y'all tagged it?		6663421286838886	http://twitter.

Figure 9: Our coding sheet - tweet text with metadata

The in_reply_to_status_id_str allows us to identify conversational threads. Reply tweets carry in this field the numerical ID of the original tweet. If a number is present in this field, then the coder knows the tweet's message has antecedents. If there is more than one tweet within a conversational thread, the in_reply_to_status_id_str will refer to the tweet that was being directly replied to.

Even more important for our purposes is the status_url. This is a permalink to the tweet in its original form on the Twitter site. By clicking on the status_url, a coder is able to see not only where the tweet is placed within a conversational thread but also any embedded media including pictures or videos.

Figure 9 shows a tweet as it is displayed in our coding sheet with its text and three fields of metadata. Without additional contextual elements, it is unclear what is meant by tagging and what object might have been tagged. The fact that there is a number in the in_reply_to_status_id_str (the third column from the left) tells us that this tweet is a reply. Clicking on the status_url, we are able to see the original

Figure 10: After following the status_url

tweet (figure 10). It becomes clear that the user was asking whether the Monterey Bay Aquarium had placed a tracker on a turtle they released into the wild.

In addition to these two pieces of metadata we've included in our coding sheet, we also thought we were losing a lot of context by not having access embedded media. The URL for any embedded images can be found under the media_url object. In the default TAGS spreadsheet, the media_url is listed along with other pieces of metadata compiled under the entities_str column. In order to pull the media_url and display it as a picture within our coding sheet, we wrote the following function in our Google Sheet:

```
=IMAGE(IFERROR(REGEXEXTRACT(R2,"media_url.:.(http.*?).,"),""))
```

This function enters the entities_str cell (column R in the default TAGS spreadsheet), returns any URL found after the string media_url":", and displays it using the image formula. When there is an embedded image present, this function will display it in the cell.

This function adds an important layer of context, but it is imperfect. First, only embedded images are captured by the media_url entity, meaning that this function will not pick up embedded videos or audio. Second, media displayed in Twitter Cards are not registered in the media_url entity. For instance, the following tweet landed in our dataset as hoodline.com/2015/12/cal-ac… @RebaFay @calacademy. After clicking on the status_url, we saw the following tweet shown in figure 11 that includes both an embedded image and teaser text for an article.

Figure 11: Tweet with Twitter Card below

Despite these cases, we still believe this to be an important

contextual layer. We began our project downloading data as .csv files from the TAGS Google Sheet and importing them into Excel for coding. It seemed more powerful and fitting for the serious work of research. The ability to embed a remotely hosted image in a cell though convinced us to begin coding within the Google Sheet itself.

The one other piece of metadata we display in our coding sheet is the from_user column, which allows us to distinguish tweets sent from the science institutions that are the focus for our research. Figure 12 shows what our coding sheet looks like with these metadata, and the tweet text.

Figure 12: Embedded images displayed in the coding sheet

WHAT IS YOUR UNIT OF ANALYSIS?

As we have seen through the lenses of cohesion, situationality, and intertextuality, constructing meaning from tweets can be challenging. Referencing the text of the tweet alone might be sufficient for more structural or quantitative analyses, but it has been our experience that awareness of these contextual elements is necessary in order to apply many of our thematic codes.

This leads us to the question of defining units of analysis. In the literature we have read, very few authors identify what exactly coders analyzed. The assumption seems to be that unless otherwise stated, the tweet is the unit of analysis. That

is one issue we hope to problematize and thus open avenues for future work.

The question of what elements to use in coding has been a struggle throughout much of our project. We started coding everything including the contents of external websites linked in tweets. This was prohibitively slow and problematic in many ways. We then tried to code from just the text of the tweet and found often we couldn't understand what it was saying.

These experiences led us to the decision that coders should first attempt to code a tweet based on what they have in their coding sheet—the text of the tweet, the sender, and embedded media. If they are unable to understand the meaning from these elements, they may click on the status_url, taking them to the tweet as it was originally presented on the Twitter website. This allows them to see any additionally embedded elements such as videos, content from Twitter Cards or retweets, and related conversational threads. Finally, if they are still unable to understand the meaning, they should click through any links and view the first screen of information.

We acknowledge that this approach opens up the possibility of inconsistency between what different coders might see and how they might interpret content, but we believe that it is the best balance between insisting that coders click on every link, which would have made the time spent on coding prohibitive, or trying to write top-level codes that could be applied without referencing contextual elements. We tried both of these options and found them to be less feasible than our solution.

As we work to develop new digitized research methods that support the exploration of online data such as Twitter content (Giles, Stommel, Paulus, Lester, Reed, 2015), it is questions like these that we will need to address.

REFERENCES

Barron, B. (2006). Interest and self-sustained learning as catalysts of development: A learning ecology perspective. *Human Development*, 49(4), 193-224.

danah boyd. (2010). "Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications." In Papacharissi, Zizi (Ed.) *Networked Self: Identity, Community, and Culture on Social Network Sites* (ed.), pp. 39-58.

Bruns, A., & Liang, Y. E. (2012). Tools and methods for capturing Twitter data during natural disasters. *First Monday*, 17(4).

Bruns, Axel, and Moe, Hallvard. (2014). "Structural layers of communication on Twitter." In K. Weller, & C. Puschmann (Eds.). *Twitter and Society*. New York: Peter Lang. Twitter and society 89: 15-28.

Beaugrande, R. D. R., DRESSLER, W. (1981). Introduction to Text Linguistics. London and New York: Longman Paperback.

Dwoskin, Elizabeth. (2014, Jun 6). In a Single Tweet, as Many Pieces of Metadata as There Are Characters. *Wall Street Journal*. Retrieved from: http://blogs.wsj.com/digits/2014/06/06/in-a-single-tweet-as-many-pieces-of-metadata-as-there-are-characters/ On March 18, 2016

Gallezot, G. (2013). Tweets & Science. In Pelissier, N & Gallezot, G. (Eds.). *Twitter Un Monde En Tout Petit*, 233-240.

Giles, D., Stommel, W., Paulus, T., Lester, J., & Reed, D. (2015). Microanalysis of online data: The methodological development of "digital CA". Discourse, Context & Media, 7, 45-51.

Gleason, B. (2013). #Occupy Wall Street: Exploring informal learning about a social movement on Twitter. American Behavioral Scientist, July 2013 vol. 57 no. 7 966-982

Halavais, A. (2013). Structure of Twitter: Social and technical. Twitter and society, 29-42.

Han, B., & Baldwin, T. (2011, June). Lexical normalisation of short text messages: Makn sens a #twitter. In Proceedings of the 49th

Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (pp. 368-378). Association for Computational Linguistics.

Honeycutt, C., & Herring, S. C. (2009, January). Beyond microblogging: Conversation and collaboration via Twitter. In System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on (pp. 1-10). IEEE.

Roomann-Kurrik, Arne (2012, June 14). Twitter Cards. *Twitter Blog*. Retrieved from: https://blog.twitter.com/2012/twitter-cards

Schmidt, J. H. (2014). Twitter and the rise of personal publics. Twitter and society, 3-14.

Shu, Catherine. (2015, April 6). Twitter Officially Launches Its "Retweet With Comment" Feature. *TechCrunch*. Retrieved from http://techcrunch.com/2015/04/06/retweetception/

St. John, M., & Perry, D. (1993). A framework for evaluation and research: Science, infrastructure and relationships. Museum visitor studies in the 90s, 59-66.

Twitter. (2016). What are replies and mentions?. *Twitter Help Center*. Retrieved from http://support.twitter.com/articles/14023 on March 18, 2016

Twitter. (2016). Summary Card. *Twitter Developers*. Retrieved from: https://dev.twitter.com/cards/types/summary on March 18, 2016.

Williams, S. A., Terras, M. M., & Warwick, C. (2013). What do people study when they study Twitter? Classifying Twitter-related academic papers. Journal of Documentation, 69(3), 384-410.

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. New media & society, 13(5), 788-806.

Zimmer, M., & Proferes, N. J. (2014). A topology of Twitter research: Disciplines, methods, and ethics. Aslib Journal of Information Management, 66(3), 250-261. Chicago