Data Craft: The Manipulation of Social Media Metadata

Amelia Acker

Data&Society

Contents

Amelia Acker; Assistant Professor, School of Information, University of Texas at Austin; PhD 2014, Information Studies, University of California, Los Angeles.

This report is published under Data & Society's Media Manipulation research initiative; for more information on the initiative, including focus areas, researchers, and funders, please visit
→ datasociety.net/research/media-manipulation

Executive Summary

This report describes how manipulators use data craft to create disinformation with falsified metadata, specifically platform activity signals. These data about engagement activities can be read by machine-learning algorithms, by platforms, and by humans. Manipulators are getting craftier at evading moderation efforts built upon these metadata categories by using platform features in unexpected ways. This report argues that social media metadata can be read as contextual evidence of manipulation in platforms. Reading metadata as a method to validate or dispute social media data can help us understand the craftiness of media manipulators. And understanding media manipulators can help pressure platforms to do better in their efforts at challenging falsified content.

Defining metadata can be hard, even for those who use it most. This is because the designation of "metadata" can change depending on who is using the data in question and for what purposes. A working definition of metadata is the names that represent aggregated data. Once data are collected, they can be assembled, classified, and organized into structures with these names. People are able to develop meaning, create claims, make decisions, and create evidence with data once it is represented in aggregate with metadata.

Manipulating metadata can be seen as a skill set, a kind of data craftwork that plays with the features and automated operations of platforms. I call this "data craft": practices that create, rely on, or even play with the proliferation of data on social media by engaging with new computational and algorithmic mechanisms of organization and classification.

We need to develop methods of reading when, where, and how manipulators leverage metadata in platforms. These methods need to account for the possibilities of data craft, which are often skillful, targeted, and organized around common fault lines in platform features. This report includes three case studies of social media metadata manipulation: politicians' accounts on Instagram, official U.S. government Twitter accounts, and the Facebook ads purchased by the Russian-based Internet Research Agency.

Based on these examples, this report argues that reading metadata can help us more fully understand the craft of data work and the many roles of metadata in platforms. It provides avenues for identifying vulnerabilities and for pressuring platforms to do better. It points to some open questions for the future of what web archives of social media data can teach us and what their status will be in the future of disinformation studies.

Introduction

Online manipulators have become adept at using the features of social media platforms to spread disinformation and influence public discourse. Researchers, journalists, publishers, and advertising agencies have known for some time that platform features in social activity streams can be gamed through a variety of techniques. For example, manipulators can generate clicks and fake engagement through astroturfing and botnets, which in turn can generate more reshares, likes, and engagement (Confessore et al. 2018; Keller 2018). Some manipulators have gone so far as to create an international marketplace of "follower factories" and "click farms" that promote celebrity profiles, create fake reviews, sell followers and views, or promote content to sell stuff to users in their personalized feeds. Most platform companies prohibit such manipulation and frequently respond by locking down, deleting, or kicking off the offending user accounts. Still, since the 2016 US presidential elections, a new kind of platform manipulation has become apparent that's harder to track and stop: politically motivated disinformation campaigns (Chen 2015). Political disinformation campaigns attempt to influence civil discourse, erode democracy with mistrust, meddle with elections, and even attack the public sphere by hijacking the platforms where more than half of Americans get their news (Shearer and Gottfried 2017). Unlike older commercial examples, these political manipulations fly under the radar of automated moderation efforts.

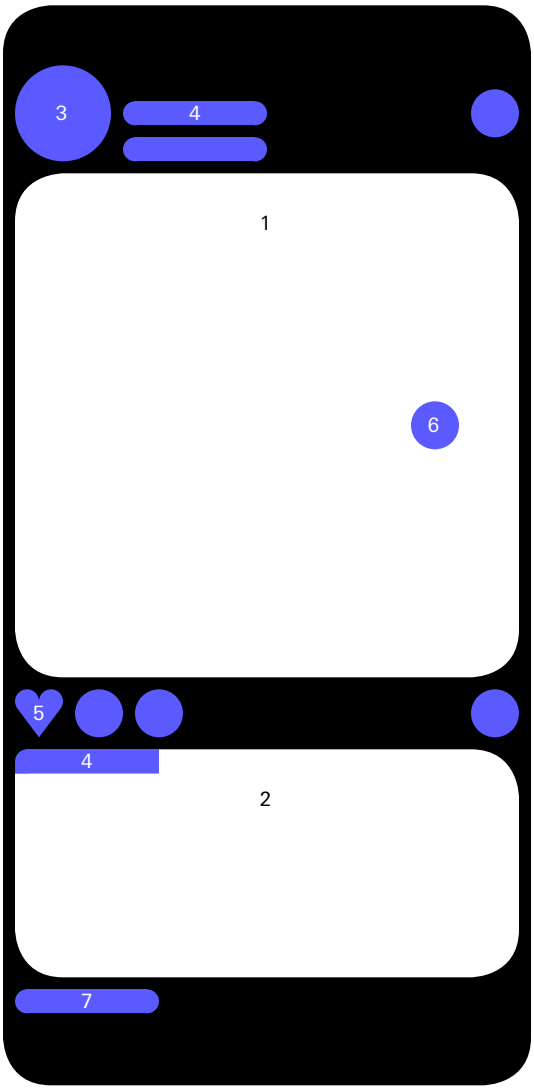
For several years, platforms have sought to automate the moderation of prohibited content with "bot sweeps" and the mass deletion of fake accounts. But automating the moderation of politically motivated manipulation has proven uniquely difficult, even when platforms have stockpiles of user data with which to train automated systems. Manipulators are getting *craftier* at faking what looks like authentic behavior on social media. In 2018 and the run up to US midterm elections, platforms like Facebook, Twitter, and YouTube have redoubled their efforts to combat disinformation campaigns, and yet, each have continued to offer services that manipulators can exploit (Dwoskin and Romm 2018; Micheal 2018). For example, by focusing on spammy bots that post incessantly, Twitter has also begun to inadvertently sweep up real people that tweet hundreds of times a day (Burnett 2018). As part of identifying false pages and profiles with links to the Russia-based manipulation group, the Internet Research Agency, Facebook has also deleted a number of event pages by legitimate community activists planning upcoming protests (Schulberg and Blumenthal 2018). For users and platforms alike, it is getting harder to discern "real" users and authentic account activities from fake, spammy, and malicious manipulations. Fig. 1 → p. 5

Whether real or malicious, all user activities are classified with *metadata* by platforms. For digital content, metadata is like the nutritional facts label on packaged food. Metadata make up the structures, stan-

Fig. 1

Differentiating Content and Metadata

This figure of a social media post highlights how content and metadata are defined in the report. In addition to the metadata available through public platform interfaces, it also captures a number of relevant metadata fields only available through approved API access.



Content	1	Photo
	2	Post
Metadata	3	Account Image
	4	Account Name
	5	Number of Likes
	6	Tag
	7	Date of Post
Additional Metadata from the API	:	*location_id* A place associated with this post.
	:	*profile_views* The number of times the account has been viewed by other users.
	:	*created_time* The creation time of the post, down to the second.
	:	*source* The application used to create the post.

dards, and tools for naming data by labeling how a digital object is generated, stored, and can be circulated in networked communication systems. Without metadata, finding or accessing information in database-driven systems is nearly impossible — it's like trying to find a house on a map without an address or a cross street. Metadata are the sign posts for transmission, access, and retrieval of information. Metadata categories can be applied flexibly to many different purposes to find, retrieve, or even provide new paths to accessing content in platforms and across the internet.[1] Metadata can be useful for quickly seeing which social media accounts have the most followers, which app store games have been downloaded the most, or how many times a YouTube video has been watched and shared. Metadata can even be used to target promoted ads on Facebook or restrict the options available on a dating platform.

This report describes how manipulators use *data craft* to create disinformation with falsified metadata, specifically with platform activity signals. Platform activity signals include a range of social media metadata: username, profile handle, bio field, dates of posted photos, followers and following counts, hearts on posts, and so forth. These data *about* engagement activities can be read by machine-learning algorithms, by platforms, and by humans. Manipulators are getting craftier at evading moderation efforts built upon these metadata categories by using platform features in unexpected ways.

This report argues that social media metadata *can be read as contextual evidence of manipulation* in platforms. Reading metadata as a method to validate or dispute social media data can help us understand the craftiness of media manipulators. And understanding media manipulators can help pressure platforms to do better in their efforts at challenging falsified content. To develop a method of reading metadata, this report begins by covering what metadata is, how users create it, and how it is used by platforms. Next, we discuss the nature of creating data for manipulative purposes as a kind of craftwork—a craftwork that can be read by examining metadata signals generated from adversarial techniques. After presenting a few cases of manipulation, we provide researchers with a method for reading metadata categories in their efforts at locating and interpreting disinformation tactics.

1 For librarians, database administrators, and network engineers, standardizing and wrangling metadata is a core part of providing access to information in collections, the provision of services, and transmission across networks. For more on the history, development, and types of metadata used in information infrastructures.
→ Pomerantz, Jeffrey. *Metadata*. MIT Press, 2015.

One Person's Metadata Are Another Person's Data

All collections of data rely on metadata. As a result, metadata are a core part of how we experience platforms. However, *defining* metadata can be hard, even for those who use it most. This is because the designation of "metadata" can change depending on who is using the data in question and for what purposes. A working definition of metadata is *the names that represent aggregated data.* For communication technologies that rely on networks, these names that represent different kinds of data are essential for sending and transmitting information, for access and retrieval mechanisms, and for searching and finding data after it's been created. Once data are collected, they can be assembled, classified, and organized into structures with these names. People are able to develop meaning, create claims, make decisions, and create evidence with data once it is represented in aggregate with metadata. This working definition of metadata relies on considering the representational problems with accurately naming data as it is being collected and used in different aggregation contexts. Some information scientists distinguish between data and metadata by examining the contexts of creation, collection, and use (Borgman 2015). But defining metadata by context is not always self-evident or unquestionable, because collections of aggregated data may change as they are used over time (Boellstorff 2013). Indeed, the line between data and metadata can be blurred by any number of changes —when those in control of data grant new access to it, change the terms of its governance, or imbue it with new status or meaning. Simply put, "One person's metadata are another person's data" (Mayernik and Acker 2018).

The way platforms define the boundary between data and metadata isn't just an intellectual exercise; it has real stakes for a range of activities now supported by apps and internet connected platforms. Perhaps the most infamous example of exploiting the relationship between data and metadata is the Cambridge Analytica breach of Facebook user data. In 2013, psychology researcher Aleksandr Kogan released a Facebook app called "This Is Your Digital Life" that collected information from a user's Facebook profile. This type of direct data collection is typical of many Facebook apps. What made Kogan's app different was that it also collected data on app-users' friends. At the time, Facebook considered data about friends as part of the metadata of a user profile (Cadwalladr and Graham-Harrison 2018).[2] Kogan was able to exploit this designation and collect data on millions of users who never interacted with his app.

Kogan and Cambridge Analytica were able to leverage this data and metadata arrangement because of a complicated set of decisions made by Facebook about access, APIs, releasing users' information, and

2 Kogan's app paid approximately 270,000 Facebook users a small fee to take a personality test for research purposes and as part of their participation in the study, and participants agreed to have their personal Facebook data collected for academic use, which included metadata about their friends such as birth dates or political party affiliations.

securing consent. A few years earlier, in April of 2010, Facebook announced the Open Graph platform, which would allow developers to create third-party apps on Facebook (like games, plugins, chatbots, or personality quizzes) (Iskold 2010).[3] The Open Graph platform API also allowed third-party developers, from data brokers to researchers like Kogan, to request access permissions to gather personal information from users through plugins. From the time that Open Graph was announced in 2010 until a developers' update in 2014, the "user data" available to developers also included some users' friends' metadata (Hern 2018). Eventually, the developers' terms of service ("Facebook Platform Policy") was updated and limited access to data from friends (Facebook n.d.). Third parties were no longer allowed to gather metadata from friends, now considered personally identifiable information by Facebook, without securing permission first. However, the developers' Platform Policy was not initially retroactively enforced for third parties like Kogan who had built apps that collected Facebook user data before the 2014 update.[4] Developers' policies, terms of use, and the enforcement terms all rely on meaningful differences between metadata "about" my friends and data "from" my friends that were not clear to users, and which allowed a third-party actor like Kogan to collect large amounts of personally identifiable information from users' friends. As a result, they were able to exploit the collection status and accessibility of users' metadata about friends, combine it with more demographic data, and create an aggregate database of psychographic profiles of users who did not use or consent to data collection from Kogan's Facebook app.

In addition to accessing collections of data, the difference between data and metadata also concerns where and how the data is created. While it may be easy to think of metadata as hidden, machine-driven, and logged automatically without our control, not all metadata is generated behind the scenes. Users themselves create much unstructured, descriptive metadata through a platform's user interface. I call this "user generated context": the metadata that people create through the everyday use of platforms—the comments they write in response to content, the "likes" they leave, the followers, the time stamps, the usernames. This metadata created by users themselves is meant to be read by others as part of the experience of interacting with platform content—comments, likes, hearts, views, or retweets; all of these are metadata. These engagement activities are layers of user generated context that, when aggregated by platforms, reveal new insights about how people communicate.

As we will see, it is the metadata from user generated context that is often gamed, hacked, and falsified by manipulators. Manipulators pay special attention to this type of metadata as they create *noisy data*, intentionally falsified data that is made to "look real." These layers of context are data traces: usernames, date and time of posting, follower counts and connections, likes or shares, and hashtags or location tags. When taken together as metadata, these labels can provide readers with clues as to how messages have been produced. In this report, I show how these social media metadata can also reveal unique behavioral signatures about an account, how platform affordances impact the possibilities of content circulation, and how people can evaluate their authenticity.

3 It's worth noting that the Open Graph API has led the way for platform APIs for many social, mobile, and video platforms. In addition to Apple's iOS and Android's developers' kits for mobile apps, this Facebook API led to a transformation in ad technology, developer tools, and new possibilities for research because of a new data access regime (replacing web scraping techniques and creating a new culture of access through APIs that can be updated and rolled back by platforms).

4 The FTC opened an investigation on Facebook's user data practices because of this reason after the Cambridge Analytica breach was reported.

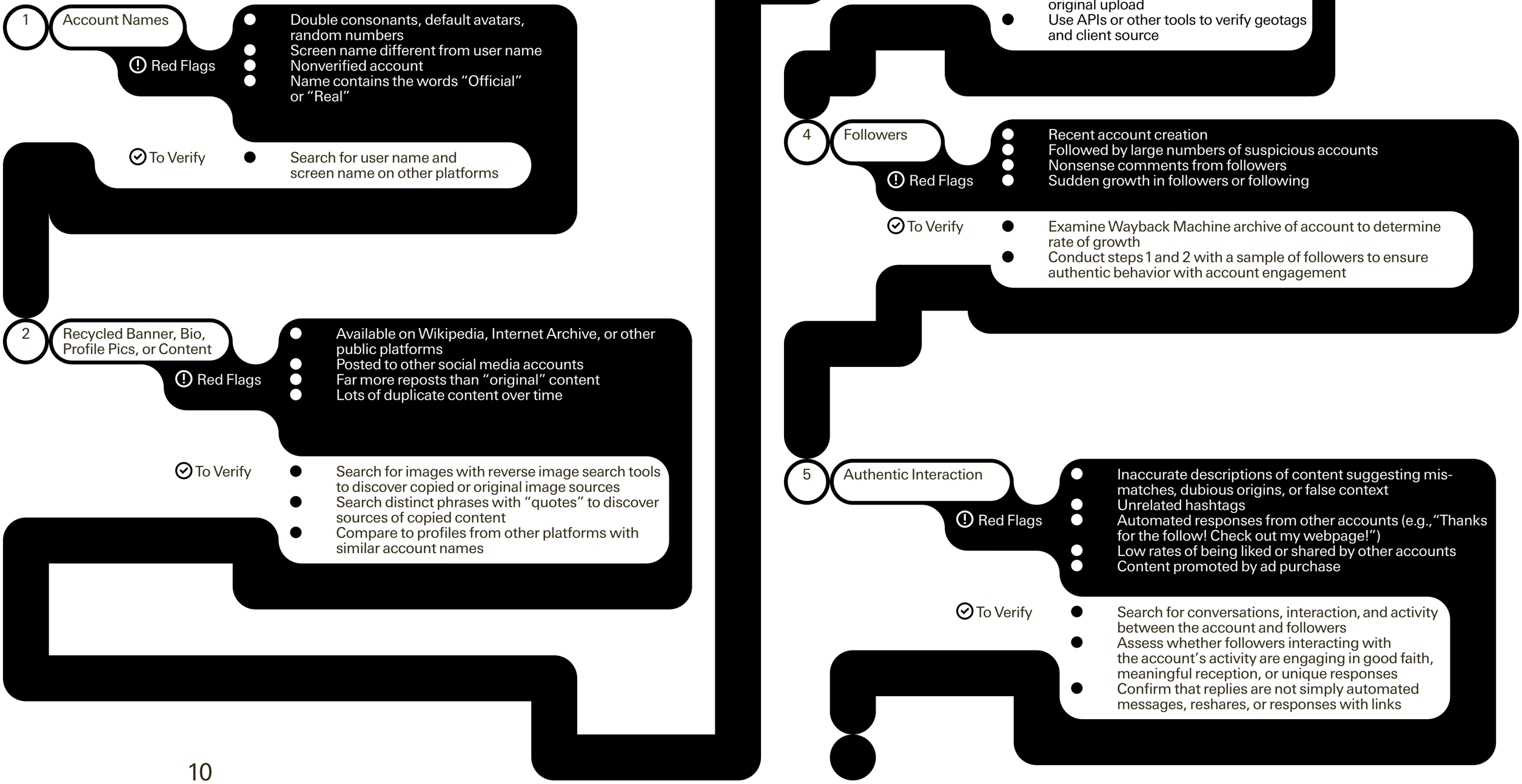
Social media researchers already study rumors, hoaxes, and disinformation through metadata (Shu et al. 2017; Starbird et al. 2018). Recently, computer scientists and security researchers who study the behavioral traits of social media users have found that patterns, or "signatures," embedded within metadata are just as unique in identifying account users as the content of social media they post (Perez, Musolesi, and Stringhini 2018). Despite plenty of research that uses social media metadata to examine the power of platforms in society, there are few examples of disinformation studies that consider manipulators' metadata strategies. When manipulators imitate authentic human behavior with fake context, they must get closer to "understanding opacity in machine-learning algorithms" (Burrell 2016). As such, manipulating metadata can be seen as a skill set, a kind of data craftwork that plays with the features and automated operations of platforms. I call this the "data craft" of disinformation, the intentional manipulation of metadata to appear authentic to both algorithmic systems and human users.

"Craft" designates work that is supplemental, material, and skillful (Adamson 2007). *Data craft* is a collection of practices that create, rely on, or even play with the proliferation of data on social media by engaging with new computational and algorithmic mechanisms of organization and classification. Data craft is one version of what Gabriella Coleman calls the "interplay between craft and craftiness" of hacking (Coleman 2016, 164). It's clear that some manipulators are craftier than others; some clumsy manipulators leave spammy signals or other incriminating digital fingerprints. As the Russian election meddling investigation has shown, some fingerprints—paying for ads in rubles, or geolocation tags of fake news from Macedonia—may only be accessible to platform engineers and those with access to APIs (Isaac and Wakabayashi 2017; Shane 2018). In many cases, the data craftwork of politically motivated manipulators has outmaneuvered automated moderation tools, at least for a time. Where signals of context get overlooked by automated disinformation efforts, *reading metadata* can help researchers, journalists, and activists concerned with adversarial tactics aimed at disinformation to locate, identify, and evaluate them.Fig.2 →p.10

Fig. 2

Reading Metadata

The chart captures a step by step process for reading metadata from social media content. The goal for each step is to evaluate different types of "red flags"—characteristics which can, when taken together, indicate likely manipulation and coordinated inauthentic behavior. None of these red flags can be interpreted as concrete evidence on their own. However, when taken together all of the following metadata categories—including interaction between other accounts—allows readers, researchers, and users to see the traces of manipulative data craft. By examining the interaction between accounts and their followers, steps 4 and 5 allow readers to locate evidence of manipulation and disinformation resulting from coordinated engagement strategies that generate inauthentic behavior.



How to Spot Metadata Manipulation

We need to develop methods of reading when, where, and how manipulators leverage metadata. These methods need to account for the possibilities of data craft, which are often skillful, targeted, and organized around common fault lines in platform features. This section uses three case studies to discuss several ways to identify, read, and locate authentic sources for metadata in cases of manipulation. At the end of each case is a specific list of tips for researchers looking to read metadata.

Babin on Instagram: Mimicking Legitimacy

We can illustrate the difficulty of identifying manipulation with an example from Instagram. Consider the similarities between these two profiles Fig.3 → p.13 — both of which claim to be the "official" account for Representative Brian Babin.

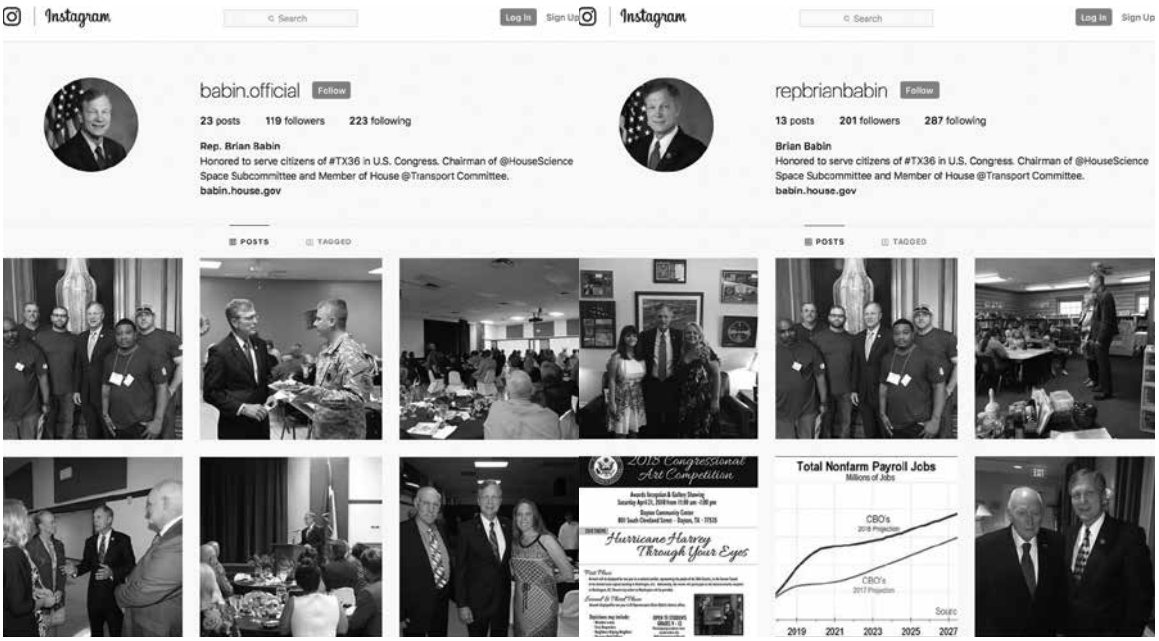
Representative Babin has served Texas' 36th congressional district since 2015 and is up for reelection in 2018. Which of these accounts is his? The babin.official account has 23 posts, 119 followers, and is following 223 users. The repbrianbabin account has 13 posts, 201 followers, and 287 followed users. Both accounts have similar profile pictures, nearly identical text in the bio field, and fewer than 300 followers. Neither account bears the blue check badge that indicates it's been authenticated by Instagram as a public figure. Though, one important detail *not* visible on Instagram: Representative Babin has verified Facebook and Twitter accounts, each with the handle: "repbrianbabin."

At first glance, babin.official and repbrianbabin both appear to be authentic Instagram accounts, but do *legitimate* users run them both? Is one of these accounts an example of coordinated inauthentic behavior, the kind of behavior that violates Instagram's terms of use?[5] Based on a simple comparison of public metadata, these two accounts are indistinguishable. Short of contacting Babin's office, we can't easily distinguish between them.

How can researchers explain the Babin accounts to us? Is one more suspicious, fake, or inauthentic than the other? It takes a closer reading of metadata to answer these questions. When considered closely (post by post), the babin.official account appears to be at least more *suspicious*, if not wholly inauthentic: the posts have no descriptive captions, few posts have comments and likes from followers, and duplicate posts (posts *also* appearing on the repbrianbabin account) are published with newer time stamps, often one or two days after they appear in the repbrianbabin feed. These signals are early indicators of how context can be mimicked, gamed, or falsified. What does this rather young and seemingly benign example tell us about how social platforms can be

5 Instagram is a Facebook product. As part of Instagram's terms of use, account holders must comply with Facebook's Community Standards.

Fig.3 Screengrabs *babin.official* and *repbrianbabin* accounts. Compare each account's username, Instagram handle, bio, and follower counts.



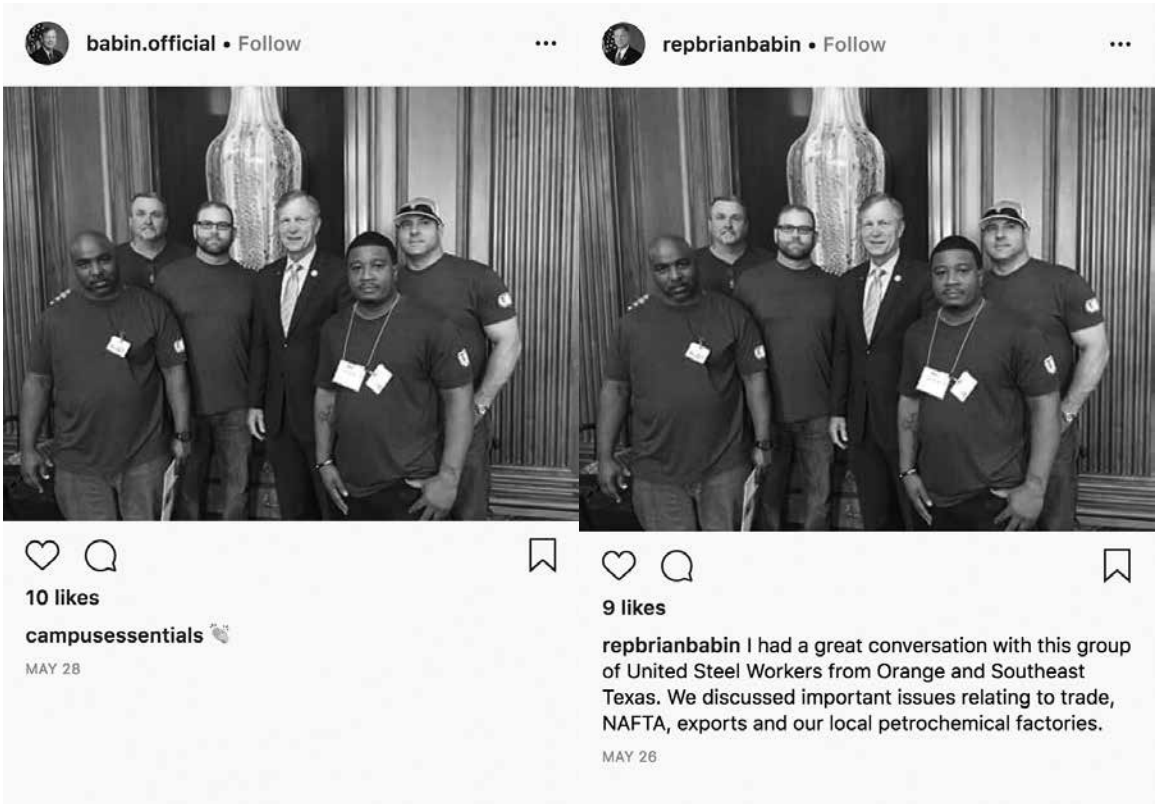
used to spread disinformation? Is babin.official malicious with the intent to deceive users? Possibly. The account could be malicious, or it could just be a way to get followers and engagement for another account or hashtag. There are many possible explanations: it could be spam; it could be a fraudulent scam to solicit donations from constituents; it could even be an early example of election meddling; or it could be the product of a new social media intern on the campaign.

The Babin Instagram accounts illustrate exactly why automated moderation cannot always spot inauthentic activity. If you aggregate each account's metadata, the two may seem nearly indistinguishable: similar follower counts, similar numbers of posts. But if you examine a few of the posts from each account individually, examine the lists of followers and following accounts, and read for the number of "hearts," the amount of comments, and the descriptions of photos, you begin to get a feel for which is the *legitimate* Babin Instagram account. Read carefully, babin.official is far less trustworthy than @repbrianbabin. The @repbrianbabin account, despite having fewer photos than babin.official, has been "tagged" by other legitimate users (Representative Phil Roe and a former staffer). Photo posts from repbrianbabin often include a description of the event; many appear to be taken with a cell phone. And the content of the images varies: several are snapshots of Babin's wife, some are screengrabs from reports and slide decks.[6] The time stamps for the photos duplicated between the accounts also tell a story. These photos always show up first on the repbrianbabin account, before being reposted on babin.official a few days later. Finally, none of the photos posted on babin.official have captions. Fig.4 →p.14

But even with this level of mimicry, it remains to be seen whether babin.official is a malicious account. Still, it does illustrate the kinds of adversarial tactics disinformation campaigns can take to make false accounts look legitimate. Despite having the early markers of a spoof account, @babin.official is at low risk of being flagged by automated

6 It is possible for other accounts to mistakenly tag the wrong Babin account, as in the two tags from babin.official on 8.8.2018. This is a tactic for gaining credibility from other accounts.

Fig.4 Identical photograph posts with different context, one posted two days after the other; repbrianbabin's post from May 26, 2018, includes a caption and no comments, babin.official's post from May 28, 2018, has no caption and has one emoji comment from another account called campusessentials. → 7



moderation techniques, because it has very few followers and followed accounts. Having more engagement, for example tens or hundreds of copied images, or hundreds of thousands of followers, could also be overlooked as authentic signals because more content and followers indicates active use of the platform.

Side by side, the two Babin accounts help us explain the larger problem of forgeries and impersonations across social media. Other impersonation techniques include "sock puppet" or "deep cover" accounts. A sock puppet account is any account a user creates with a false identity — whether for satire or deception. Misleading sock puppet accounts are often used to evade platform bans or to stuff ballots by creating multiple "puppet" accounts. Deep cover accounts involve developing online identities that build up a network of content and followers over time, until they are mistakenly attributed to a real person or organization. If users and platforms accept these signals as real, the authenticity of a forged, impersonating account is less likely to be challenged, especially if no other competing accounts exist or they have little or no activity when compared. Deep cover and spoof accounts can continue to mislead and still bring traffic, gain followers and interactions, and appear real, until another account is found that shows mimicry and its legitimacy is challenged as impersonation. In cases of simple impersonation, user generated context such as dates of creation or posting, usernames, photo tags, and account handles, can make it easier for researchers, reporters, and users to discover the truth.

7 Reference Links
→ https://www.instagram.com/p/BjPcVlcH9QM/?taken-by=repbrianbabin (posted May 26)
→ https://web.archive.org/save/https://www.instagram.com/p/BjPcVlcH9QM/?taken-by=repbrianbabin (Internet Archive capture)
→ https://www.instagram.com/p/BjUkqEkIsFj/?taken-by=babin.official (posted May 28)
→ https://web.archive.org/web/20180813194052/https://www.instagram.com/p/BjUkqEkIsFj/?taken-by=babin.official (Internet Archive capture)

Tips

- Search for accounts with the similar or same real name, username, or account handle across other platforms
- Compare profile pics and account descriptions across multiple platforms
- Identify copies of posts and their time stamps
- Reverse Google image search profile pics and account banners
- Locate absences that have not been copied or backfilled (e.g., lack of comments, description, etc.)
- Check if an account has been tagged by other verifiable platform users
- Consider parody or organizational change as explanations before concluding that something is malicious manipulation

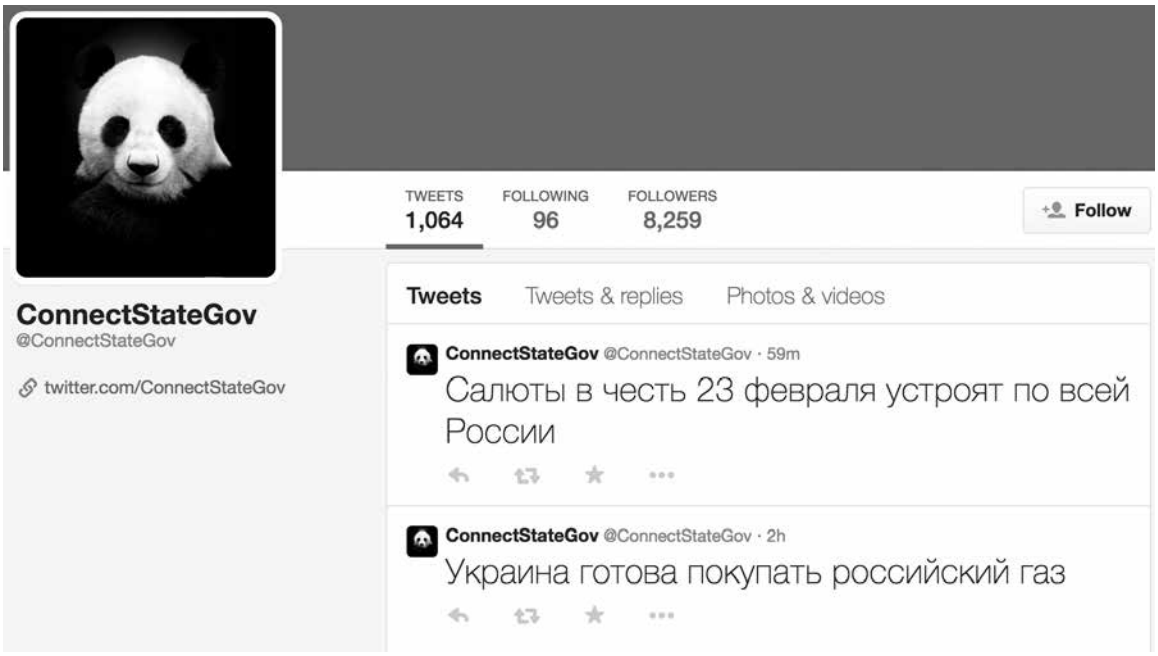
Creating Imposter Accounts: Claiming Deleted Screen Names

Examining metadata isn't always a direct line to uncovering manipulated material. Some crafty manipulation techniques involve intentionally leaving behind fingerprints to create a sheen of authenticity. For example, some manipulators are able to accomplish full account takeovers. Rather than creating a parallel impersonation account, these manipulators gain access to and control over the original, legitimate account. In these cases, account metadata may look real because of previous, legitimate online activity. Still, sometimes newer signals may appear disjointed from earlier content, such as tweeting in a new language or in a new tone of voice.

In November of 2017, Justin Littman, an archivist at George Washington University Libraries noticed that several US government accounts were tweeting in Russian (Littman 2017a). As part of his work with the Social Feed Manager (a digital preservation tool that allows researchers and archivists to gather social media data), Littman was collecting and preserving tweets from nearly 3,000 government agencies (GWU Libraries 2018). Initially, Littman consulted the US Digital Registry for confirmation that the accounts were authentic. The Registry is the official list of US government accounts across all kinds of social media platforms, including mobile apps, to confirm that these were indeed official government accounts. It is run by Digital.gov ("US Digital Registry" n.d.). According to its website, the Registry's reference database is intended "To help prevent exploitation from unofficial sources, phishing scams, or malicious entities, the US Digital Registry serves as a crowdsource resource for agencies, citizens, and developers to confirm the official status of social media and public-facing collaboration accounts."

Littman found 100 deleted accounts and 29 suspended accounts that were still listed as active, official government accounts by the Digital Registry. He had discovered a vulnerability with Twitter screen names, where imposter accounts are created by claiming the screen name of an abandoned account. Twitter's policy is to remove accounts that have no activity after six months. Therefore, if an official US account is inactive for long enough, anyone can submit a request to take over the associated screen name. Littman documented that, apparently, over 100 government agencies had abandoned their accounts between 2016 and 2017. Twitter had removed many of these for inactivity even as the US Digital Registry still listed them as active. All that remained was for manipulators to swoop in on the abandoned accounts. Fig.5 →p.16

Fig.5 Official US government account tweeting in Russian from Wayback Machine capture, October 2014. →8



To demonstrate this vulnerability, Littman acted as an adversary and "attacked" the official @USEmbassyRiyadh account. He waited for it be abandoned and then claimed the screen name himself (Littman 2017b). First, he changed his Twitter screen name to @USEmbassyRiyadh, then copied the official banner image, profile, name, location, and other metadata from the Internet Archive. Then Littman, impersonating the US Embassy in Ryadh, Saudi Arabia, tweeted a Wilford Brimley quote. He closed the loop on this attack by archiving the imposter page in the Wayback Machine's web archive. Fig.6 →p.17

Littman's experiment shows how exploiting a handful of metadata categories, like usernames, Twitter account handles, profile pics, and banners, can exploit the reserves of archive.org and other digital archives. By locating (and confirming) an outdated official registry, reanimating a screen name, gathering authentic signals from 2015, and then archiving the imposter account with the Wayback Machine, Littman's data craft reveals how ongoing, legitimate web archiving efforts of libraries, researchers, and nonprofit cultural institutions can also be exploited by manipulators.

Since this attack last year, Digital.gov has invited all agencies to review and update their accounts.[9] Littman continues to collect social media data from US government agencies and has called publicly for government agencies to take advantage of Twitter's verified status option.

Tips

- Locate date of when account was started, user joined
- Look at how many tweets/posts have been created since account start date
- View attached media (pictures, videos, links) and look for duplicates
- Consider the date of the last post or activity and if the account has been dormant

8 Sources
→ https://web.archive.org/web/20141014121748/https://twitter.com/ConnectStateGov

9 Official US government accounts are updated independently by federal workers who maintain each of the social media accounts independently. As of April 19, 2019, accounts that have not been active or updated by agencies since January 1, 2017, have since been archived by the US Digital Registry and platforms.

- If it is listed as an official account, search for the personality/institutional home page to cross-reference and confirm existence of an official account
- Search the Internet Archive's Wayback Machine for crawls of the account
- Scan multiple crawls if possible and look to see if the account was dormant or has ever been deleted or suspended
- Reverse Google image search the profile pics and banners to see if and where copies occur
- Explore followers and commenters for their authenticity to assess if they appear to be real people or bots
- Read to see if the comments are substantive and engaging with the content or if they are simply reactions or emoji

Fig. 6 US Embassy Riyadh tweet, "It's the right thing to do & a tasty way to do it." → 10

Facebook Internet Research Agency Ads

If web archives can be exploited by manipulators, they can also be used to trace how their manipulation campaigns unfold. As archivists like Littman discovered, our web archives are now filled with examples of manipulation that were, at first, overlooked by platforms. Currently, when these traces are discovered, they are *disappeared* from platforms. The imperative for web archives then, is to collect social media data apart from platforms, so that it can be used by researchers, historians, journalists, and citizens. This preservation mandate becomes even more apparent when you consider how few collections of disinformation campaign data exist beyond those released as part of congressional hearings in 2018.

In May of 2018, the US House Intelligence Committee published 3,517 Facebook and Instagram ads that were purchased by the Internet Research Agency (IRA), a Russian propaganda firm ("Exposing Russia's

10 Source
→ https://web.archive.org/web/20171107054431/https://twitter.com/USEmbassyRiyadhyoutube.com %2Fwatch%3Fv%3DGij0RgShO%7C&id__exact=789

Fig.7 Promoted ad #789, "Police is not above the law!"

Effort to Sow Discord Online: The Internet Research Agency and Advertisements | US House of Representatives" n.d.). Facebook originally shared the ads with Congress as part of the Committee's open hearing on social media companies and election meddling. The ads themselves were transmitted in zipped PDF files and redacted by Facebook to protect users' personally identifiable information. The release did *not* include the 80,000 organic posts shared on Facebook by the IRA, but the Committee hopes to make the organic content publicly available in the future (Lapowsky 2018).

By providing the data as a PDF, Facebook employed its own version of data craft — leveraging their knowledge of how data and metadata gets processed to make it difficult for other parties to work with the data to exhume and analyze trends. The PDF file format is one of the hardest digital formats to extract structured data from. However, soon after the release of the IRA ads, digital archivist Ed Summers created software to extract images and metadata from the PDFs and output them into easy-to-read JSON files (Summers [2018] 2018). Working off of Summers' program, developer Simon Willison wrote software that converts the JSON files into a searchable database now available on the web ("Russian Internet Research Agency Facebook Ads: Russian-Ads" n.d.): https://russian-ira-facebook-ads.datasettes.com/. The ads database from Summers and Willison provides a closer look at how user generated context, politicized content from manipulators, and promoted demographic targeting categories powered by Facebook's powerful ad technology were all leveraged for the IRA's influence campaign on the 2016 Presidential elections. We can examine this database to see exactly which metadata fields were central to the IRA's data craft.

Here is a promoted post, from the page Williams&Kalvin, which features an ad for a Youtube video titled, "Police is not above the law!" Fig.8

The Williams&Kalvin account was actually run by Russian manipulators and posted content that appeared across YouTube, Facebook, and Twitter. The IRA ads database reveals that Williams&Kalvin would make original YouTube videos and then buy ads to promote them on Facebook as in ad #789, "Police is not above the law!" The account had very little interaction with other users, but instead pushed out frequent video content and bought promotional ads on Facebook that would link back to their now banned YouTube page. Initially the account posted content about police brutality and racism and was consistent with some Black Lives Matter social media content. After building a following, and closer to the 2016 election, the account frequently posted anti-Clinton content containing conspiracy theories. The Williams&Kalvin account

Fig.8 Database entry for Ad 789 → 11

id 789

img

targeting location:United States, age:18–54, language:English (UK), language:English (US), placements:News Feed on desktop computers, placements:News Feed on mobile devices, accessing_facebook_on:Wi-Fi, people_who_match:interests:BlackNews.com, people_who_match:interests:HuffPost Politics, people_who_match:interests:HuffPost Black Voices, and_must_also_match:behaviors:African American (US)

impressions 3067

clicks 172

url https://www.youtube.com/watch?v=Gij0RgShO|

text Where is the justice? Our brothers and sisters are being cruelly killed by the so-called police every day and ourjudicial system is absolutely blind. We are all Americans, but why does our corrupt Government differ black and white people? We want the same attitude! I don't want to be scared of living in my country! They will never shut me up! https://www.youtube.com/watch?v=Gij0RgShO| Where is the justice? Our brothers and sisters are being cruelly killed by the so-called police every day and ourjudicial system is absolutely blind. We are... Police is not above the law!

spend_usd 16.0

spend_amount 1000.00

spend_currency RUB

created 2016-01-05T02:04:48-08:00

ended 2016-01-07T02:03:00-08:00

Advanced export JSON shape: ○ Default ○ Array

CSV Options: ○ Download File ○ Export CSV

```
CREATE VIEW display_ads AS
        select ads.id,
            case when image is not null then
                json_object("img_src", "https://raw.githubusercontent.com/edsu/irads/03fb4b/site/" || image, "width", 200)
            else
                "no image"
            end as img,
            json_group_array(
                json_object(
                    "label", targets.name,
                    "href", "/russian-ads/display_ads?_target="
                        || urllib_quote_plus(targets.id)
                )
            ) as targeting,
            ads.impressions, ads.clicks, ads.url, ads.text,
            cast(case
                when ads.spend_currency == "RUB" then ads.spend_amount * 0.016
                else ads.spend_amount
            end as float) as spend_usd,
            ads.spend_amount, ads.spend_currency,
            ads.created, ads.ended
        from ads
            join ad_targets on ads.id = ad_targets.ad_id
            join targets on ad_targets.target_id = targets.id
        group by ads.id
        order by ads.id
```

11 → https://russian-ira-facebook-ads.datasettes.com/russian-ads-919cbfd/display_ads?_search=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3DGij0RgShO%7C&id__exact=789

even appears to have done A/B testing by posting similar ads with different messages to see which versions increased engagement and followers (Lapowsky 2018).

By comparing Figure 5 and Figure 6, we can see that the metadata from the database includes several fields that users could see on the platform. Figure 6 includes the text captioning the video and the ad's default "landing page," in this case a link to the video on YouTube. However, the bulk of the metadata found in Figure 6, such as the ad targeting fields, the impressions and clicks counts, and the cost of the ad in its original currency reveals quite a bit more about the data craft used by the IRA. When buying ad #789 to promote to Facebook users, the Williams&Kalvin account selected two types of racial targeting "segments": interest matches and ethnic affinity categories. Interest matches are targeting options based on content that Facebook users engage with directly, like, share, or comment upon within the platform. These interests can include many proxies for race and ethnicity, such as sharing "HuffPost Black Voices" articles or liking *BlackNews.com*. Ethnic affinity categories are based on data gathered from outside the Facebook platform. Facebook purchases from third-party data brokers that work with ad platforms (Angwin, Mattu, Paris 2016).

The combination of these two metadata categories have proven to be remarkably powerful in targeting promoted content to (or away from) Facebook users. In 2016 and 2017, investigative journalists at *ProPublica* discovered that internal Facebook classifiers and external ethnic affinity categories in the advertising platform could be used to prevent housing ads from being seen by African Americans or Asian Americans (possibly violating discrimination in housing laws), or to promote content to users with anti-Semitic interests (Angwin and Varner 2017). After both investigations, Facebook responded to the issue with automation — updating the ads platform so that it would disable the use of ethnic affinity marketing for particular kinds of ads and removing interests that were explicitly discriminatory (Egan 2016; Sandberg 2017). Still, even with these efforts at addressing discrimination and promoting inclusion in advertising on the platform, it is clear that ethnic affinity marketing can still be misused while flying under the radar of increased human review of automated processes of enforcement.

Over half of the ads from the IRA dataset targeted race segments such as interests in "*BlackNews.com*" or "Black Voices" or "African American." Nearly a quarter of the ads in the dataset involved the targeted segments involved policing issues such as "Police Misconduct" and "Stop Police Brutality."[12] With this material, the IRA sought to garner attention and support from groups both interested in racial equality as well as those opposed to it. By using the Facebook ads interface, which uses the metadata of users' accounts, the IRA was able to collect and sort audiences for counterintelligence operations with ease. Further, when looking at the ad targeting segments selected by Williams&Kalvin, user data, including interest in *The Huffington Post* or reading the news feed on one's mobile device, became useful as Williams&Kalvin focused their campaign. Associated interests, partisan media, and device choices can act as proxies for more specific targeting of politics, class, gender, and race.

What were the signs that the IRA ads were inauthentic? Were the manipulators crafty in their tactics? Now that we have the data, we find several obvious fingerprints (e.g., accounts with little interaction and inauthentic conversations). Further review of the account administrators, the IP addresses from where content was created, the locations of most

12 For more on the top targeted segments
→ https://russian-ira-facebook-ads.datasettes.com/russian-ads-919cbfd/top_targets

of the accounts' followers, and even in some cases (as in the Williams&Kalvin post above) the money used to buy the ad, reveals an intent to deceive Facebook users. While the IRA paid little heed to Facebook's platform terms of service, they could operate in plain sight because Facebook's moderation relies on users to flag suspicious content. Only recently has Facebook began to preemptively look for violations to their TOS with a more stringent app review process for developers (Flynn 2018).

Tips

- Use social dashboards to see account creation time and date, and average daily active posts, comparing the date of the account establishment and the posts per day
- Examine promoted posts and ads policies of platforms, research
- Consult the page administrator's user account and page, comparing the rate of posting promoted content to free content
- Examine how often content is shared (e.g., are memes or videos frequently reshared but with different captions?)

Conclusions

Platforms are in the business of creating metadata — for themselves, for the developers who build on top of the platforms, and for the data brokers that buy access to it.[13] In addition to user generated contextual metadata, we can identify and *read* many additional metadata fields by examining developer's documentation, platform policies, and terms of service. Metadata are indexes of human behavior, and like sign posts, they provide paths for us to follow. But sometimes these paths are a challenge to follow—metadata structures can be hidden, entrenched, unknown, or simply inaccurate. Reading metadata depends on platform literacy and an analytic method able to account for how data is created, how it flows through information infrastructures, and how it is vetted across the internet.

Politically motivated manipulators understand the representational problems with naming data from platform activity signals because their techniques rely on creating a gap between accurate representation of legitimate platform activity signals and falsified ones. Working within this gap is how manipulators are getting craftier and more agile at avoiding automated moderation techniques. As part of their craftiness they not only create noisy, illegitimate data to be named and aggregated with authentic data, they are also in deep dialogue with the platform moderation policies; the algorithms driven by personalization and ad technology; and those features that keep users engaged with platform content in ways that regular social media users are not. In these ways data craft is about manipulating a system to assert power over it, and in doing so it can reinforce and reveal limits, or even blind spots in platforms.

Researchers, too, can develop this craft by considering the contexts in which social media data are created, collected, and named. The craft of reading metadata involves actively toggling between contexts, going back and forth across layers of account activities, and judging intent and authentic behavior in spaces where sometimes little can be inferred from the labels individually. But when taken together, these signs reveal a broader profile. In addition to providing insight to the infrastructure that undergirds social media, reading metadata serves as a method for identifying disinformation. Perhaps most importantly, reading metadata is means for apprehending the opacity of the machine-learning algorithms that increasingly drive social media platforms and lifting the veil of corporate secrecy over how data from disinformation and fraudulent activity signals are represented, identified, and then leveraged for automated moderation techniques.

13 There's even more metadata that's created when people use mobile networks and the internet to use platforms such as the transportation layer encryption, anonymization protocols, obfuscation standards and techniques, in addition to users' metadata connected to their Internet Service Provider or their mobile phone company or handset. Not sure if it's worth elaborating, but it's a stack that is meaningful for surveillance, policing, state actors, and abiding by different international laws about internet privacy.

The IRA ads database itself points to the problem of scale for platforms attempting to combat disinformation and coordinated inauthentic behavior. The problem of scale is insurmountable and cannot be automated without significant consequences to censoring user generated content. It's a problem for researchers, journalists, or citizens interested in documenting change, too. In the current regime, once disinformation or inauthentic behavior has been identified it is deleted as a matter of procedure. And even with authentic data and metadata, the techniques for data extraction remain partial, incomplete, and subject to API rollbacks. Computer scientists and digital preservation scholars have argued that the long-term "changingness" and decay of metadata after it's been collected from platform APIs reveals a massive vulnerability in the persistence of social media (Zubiaga 2018; Walker 2017). That vulnerability is the inability to preserve the context, data, and metadata from manipulation campaigns as evidence of disinformation spread in platforms for researchers, journalists, policy makers, and historians of this moment. Fighting disinformation on social platforms then, isn't just a matter of better automation to flag inauthentic content. It's also a matter of solving the archival dilemma of providing long-term access to the data deployed in manipulation campaigns as they were represented in platforms.

In this report on reading metadata, we covered how to identify some adversarial tactics for creating noisy data, for faking legitimacy, and for targeting real users with intent to deceive. To game, falsify, or hack metadata categories requires both a crafty mentality and deep knowledge of platforms and their data. Such data craftwork reveals the inner workings of social platforms, whether for good or for ill. Reading metadata and data craftwork can also be applied to other forms of digital culture beyond tactics for political manipulation, scamming users for money, or garnering social capital with influence campaigns.

Based on these brief examples, this report has shown how reading metadata can help us more fully understand the craft of data work and the many roles of metadata in platforms. This report has also provided some avenues for identifying vulnerabilities and for pressuring platforms to do better. It has pointed to some open questions for the future of what web archives of social media data can teach us and what their status will be in the future of disinformation studies.

References

Adamson, Glenn. 2007. *Thinking through Craft*. Bloomsbury Publishing. https://www.bloomsbury.com/uk/thinking-through-craft-9781845206475/

Angwin, Julia, and Madeleine Varner. 2017. "Facebook Enabled Advertisers to Reach 'Jew Haters.'" Text/html. *ProPublica*. September 14, 2017. https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters

Boellstorff, Tom. 2013. "Making Big Data, in Theory." *First Monday* 18 (10). http://firstmonday.org/ojs/index.php/fm/article/view/4869

Borgman, Christine L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Massachusetts: The MIT Press.

Boucher, Tim. 2018. "Adversarial Social Media Tactics." *Tim Boucher* (blog). August 10, 2018. https://medium.com/@timboucher/adversarial-social-media-tactics-e8e9857fede4

Burnett, Sara. 2018. "Crackdown on 'Bots' Sweeps Up People Who Tweet Often." AP News. August 4, 2018. https://www.apnews.com/06efed5ede4d461fb2eac5b2c89e3c11

Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (1): 2053951715622512. https://doi.org/10.1177/2053951715622512

Cadwalladr, Carole, and Emma Graham-Harrison. 2018. "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach." *The Guardian*, March 17, 2018, sec. News. http://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Chen, Adrian. 2015. "The Agency." *The New York Times*, June 2, 2015, sec. Magazine. https://www.nytimes.com/2015/06/07/magazine/the-agency.html

Coleman, Gabriella. 2016. "Hacker." In *Digital Keywords: A Vocabulary of Information Society and Culture*, edited by Benjamin Peters. Princeton University Press.

Confessore, Nicholas, Gabriel J.X. Dance, Rich Harris, and Mark Hansen. 2018. "The Follower Factory." *The New York Times*, January 27, 2018, sec. Technology. https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html

Dwoskin, Elizabeth, and Tony Romm. 2018. "Facebook Says It Has Uncovered a Coordinated Disinformation Operation Ahead of the 2018 Midterm Elections." *The Washington Post*. July 31, 2018. https://www.washingtonpost.com/technology/2018/07/31/facebook-says-it-has-uncovered-coordinated-disinformation-operation-ahead-midterm-elections/

Egan, Erin. 2016. "Improving Enforcement and Promoting Diversity: Updates to Ethnic Affinity Marketing." *Facebook Newsroom* (blog). November 11, 2016. https://newsroom.fb.com/news/2016/11/updates-to-ethnic-affinity-marketing/

"Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements | US House of Representatives." n.d. Accessed August 14, 2018. https://democrats-intelligence.house.gov/social-media-content/

Facebook. n.d. "Platform Policy." Facebook for Developers. Accessed August 1, 2018. https://developers.facebook.com/policy/

Flynn, Kerry. 2018. "Facebook Quietly Pauses All New App and Bot Reviews." Mashable. March 28, 2018. https://mashable.com/2018/03/28/facebook-pauses-new-chatbots-apps-platform/

GWU Libraries. 2018. "Social Feed Manager." Social Feed Manager. 2018. https://gwu-libraries.github.io/sfm-ui/

Hern, Alex. 2018. "How to Check Whether Facebook Shared Your Data with Cambridge Analytica." *The Guardian*, April 10, 2018, sec. Technology. http://www.theguardian.com/technology/2018/apr/10/facebook-notify-users-data-harvested-cambridge-analytica

Isaac, Mike, and Daisuke Wakabayashi. 2017. "Russian Influence Reached 126 Million Through Facebook Alone." *The New York Times*, October 31, 2017, sec. Technology. https://www.nytimes.com/2017/10/30/technology/facebook-google-russia.html

Iskold, Alex. 2010. "Facebook Open Graph: The Definitive Guide for Publishers, Users and Competitors — ReadWrite." ReadWrite. April 23, 2010. https://readwrite.com/2010/04/23/facebook_open_graph_the_definitive_guide_for_publishers_users_and_competitors/

Keller, Michael H. 2018. "The Flourishing Business of Fake YouTube Views." *The New York Times*, August 11, 2018, sec. Technology. https://www.nytimes.com/interactive/2018/08/11/technology/youtube-fake-view-sellers.html

Lapowsky, Issie. 2018. "House Democrats Release 3,500 Russia-Linked Facebook Ads." *Wired*, May 10, 2018. https://www.wired.com/story/house-democrats-release-3500-russia-linked-facebook-ads/

Littman, Justin. 2017a. "Suspended U.S. Government Twitter Accounts." Social Feed Manager. November 4, 2017. https://gwu-libraries.github.io/sfm-ui/posts/2017-11-04-digital-registry

————. 2017b. "Vulnerabilities in the US Digital Registry, Twitter, and the Internet Archive." Social Feed Manager. November 6, 2017. https://gwu-libraries.github.io/sfm-ui/posts/2017-11-06-vulnerabilities

Mayernik, Matthew S., and Amelia Acker. 2018. "Tracing the Traces: The Critical Role of Metadata within Networked Communications." *Journal of the Association for Information Science and Technology* 69 (1): 177–80. https://doi.org/10.1002/asi.23927

Micheal, Casey. 2018. "If Facebook Is Removing Pages, Why Is All This Russian Material Still Up?" August 1, 2018. https://thinkprogress.org/if-facebook-is-removing-pages-why-is-all-this-russian-material-still-up-7ee4b7059b63/

Perez, Beatrice, Mirco Musolesi, and Gianluca Stringhini. 2018. "You Are Your Metadata: Identification and Obfuscation of Social Media Users Using Metadata Information." *ArXiv:1803.10133 [Cs]*, March. http://arxiv.org/abs/1803.10133

"Russian Internet Research Agency Facebook Ads: Russian-Ads." n.d. Accessed August 14, 2018. https://russian-ira-facebook-ads.datasettes.com/

Sandberg, Sheryl. 2017. "Facebook Wall Post." Facebook. September 20, 2017. https://www.facebook.com/sheryl/posts/10159255449515177

Schulberg, Jessica, and Paul Blumenthal. 2018. "How Facebook Decided Anti-Racist Activists Were Part Of A Foreign Influence Operation." *Huffington Post*, August 2, 2018, sec. Politics. https://www.huffingtonpost.com/entry/no-unite-the-right-protest-facebook-russia_us_5b632aa2e4b0de86f49efd4d

Shane, Scott. 2018. "The Fake Americans Russia Created to Influence the Election." *The New York Times*, January 20, 2018, sec. US. https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html

Shearer, Elisa, and Jeffrey Gottfried. 2017. "News Use Across Social Media Platforms 2017." *Pew Research Center's Journalism Project* (blog). September 7, 2017. http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/

Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. "Fake News Detection on Social Media: A Data Mining Perspective." *SIGKDD Explor. Newsl.* 19 (1): 22–36. https://doi.org/10.1145/3137597.3137600

Starbird, Kate, Dharma Dailey, Owla Mohamed, Gina Lee, and Emma S. Spiro. 2018. "Engage Early, Correct More: How Journalists Participate in False Rumors Online During Crisis Events." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 105:1–105:12. CHI '18. New York, NY, USA: ACM. https://doi.org/10.1145/3173574.3173679

Summers, Ed. (2018) 2018. *Irads: Working with 3,517 Internet Research Agency Facebook Ads Released by Congress*. Python. https://github.com/edsu/irads

"US Digital Registry." 400AD. DigitalGov. 06:56 - -0400 400AD. /services/u-s-digital-registry/.

Walker, Shawn. 2017. "The Complexity of Collecting Digital and Social Media Data in Ephemeral Contexts." Thesis. https://digital.lib.washington.edu:443/researchworks/handle/1773/40612

Zubiaga, Arkaitz. 2018. "A Longitudinal Assessment of the Persistence of Twitter Datasets." *Journal of the Association for Information Science and Technology* 69 (8): 974–84. https://doi.org/10.1002/asi.24026

Acknowledgments

I would like to thank members of the Media Manipulation Initiative at the Data & Society Research Institute, who greatly influenced my thinking and approach to this publication. Members of this team include Kinjal Dave, Brian Friedberg, Becca Lewis, and Britt Paris. I am very grateful for the work of Alex Litel, Justin Littman, Ed Summers, Simon Willison, and Leon Yin who made each of the case studies possible. This report would not be possible without the generous comments from reviewers and colleagues. Finally, I would like to thank my thought partner Joan Donovan, my editor Patrick Davison, and the incredible staff members at Data & Society that made this research possible from the beginning.

Data & Society

Data & Society is an independent nonprofit research institute that advances new frames for understanding the implications of data-centric and automated technology. We conduct research and build the field of actors to ensure that knowledge guides debate, decision-making, and technical choices.

www.datasociety.net
@datasociety

Design & Infographics
Scott Vander Zee
(Hubertus Design)
www.hubertus-design.ch