Economic and Social Research Council

Acceptability to participants of novel data linkages, ethical issues, and the practicalities of obtaining consent

Introduction

Increasing Internet and smartphone uptake provides an opportunity for social science. A growing body of research shows that behaviours, opinions and physiological states can be measured passively through connected devices that provide network information (both social media and smartphone address book), geolocation and health data (e.g. Sugie 2018). Table 1 provides an overview of the distinctive contribution of Digital Trace Data (DTD) to social research strategy and design (adapted from Edwards *et al.* 2013). Unlike conventional methods, such as interviews and surveys, DTD provide extensive (large samples) and locomotive (in process, as opposed to snap-shot) data capture.[1] Notably, Internet search data have been used to track the spread of influenza in the US (Ginsberg *et al.* 2009) and to build psychological constructs of nations linked to GDP (Noguchi *et al.* 2014). In the UK, open source social media communications have been used to investigate the spread of hate speech following terrorist attacks (Williams & Burnap 2015), to estimate offline crime patterns (Williams *et al.* 2016, 2019), to predict election outcomes (Burnap *et al.* 2016) and to estimate happiness levels relative to environment (Seresinhe *et al.* 2019). It is now well established that DTD are useful for measuring behaviours and opinions in the 'offline' world (see Moat *et al.* 2014). Indeed, answers to some survey questions have been predicted using social media data (Murphy *et al.* 2013).

		Research Design/Data	
		Locomotive	Punctiform
Research Strategy	Intensive	E.g. Participant Observation	E.g. Qualitative interviewing
	Extensive	E.g. Digital Trace Data	E.g. Population Surveys

Table 1: Contribution of Digital Trace Data (DTD) to social research strategy and design

Despite the reported success of studies claiming to have used DTD to estimate a range of outcomes, these new forms of data are not without their limitations. Social media provide impoverished data, lacking individual level demographic details of users, a mechanism for pre-defining reliable and valid measures and representative samples of the population. This presents an issue for understanding microlevel change of key indicators, a significant goal in longitudinal research. Therefore, it is unlikely that DTD could act as a *surrogate* to conventional methods that are designed to generate more complete datasets on populations. Instead, researchers have recently considered how DTD can *augment* traditional social research designs. The ability to predict survey measures with DTD (Murphy *et al.* 2013) suggests the potential for their use in understanding the possible extent of nonresponse bias (AAPOR 2014) or for nonresponse adjustment, which has been achieved using linked administrative data such as health records (Gray *et al.* 2013, Gorman *et al.* 2014). In addition,

[1] The term 'locomotive' is used instead of 'longitudinal' to differentiate between the captured temporal resolutions (e.g. seconds instead of years).

DTD can add a new dimension to surveys, allowing for the collection of linked digital data in real-time from a variety of sources including mobile phones, websites, apps and social media. These digital sources allow for the collection of data at various levels. Table 2 (see appendix) shows examples of linking at the aggregate and individual level ex ante (A, D) and ex post (B, C, E) (Stier *et al.* 2019).

These examples of data linkage are promising, but they also demonstrate the challenges of obtaining informed consent from survey respondents to take part in the DTD element of the research process. This may be particularly challenging where the DTD element requires prolonged data collection, in either passive or active form, given the legal requirements of the General Data Protection Regulation (GDPR). Recent work conducted by the ESRC Social Data Science Lab at Cardiff University, in partnership with NatCen and the University of Essex, explored issues of informed consent, privacy and harm in relation to linking DTD (Twitter in particular) to survey data. This document outlines our findings in the context of other relevant work.

Practicalities of obtaining consent

Research shows issues in conventional social science methods, such as non-response and missing data in surveys, may be mitigated by linking to alternative sources of digital data, such as those generated on smartphones, websites and social media (AAPOR 2014). For example, researchers have postulated that unit nonresponse may be mitigated by data passively gleaned from a user's linked social media account where prior consent has been secured (Al Baghal *et al.* 2019). However, DTD, and social media data in particular, are not a panacea for the failings of surveys, and linkage is only feasible for respondents who generate digital traces via smartphones and online interactions, and who agree to have these linked to a survey. Even then, variation in technology usage and consent rates may result in sample bias. Further research is required to identify the likely extent of such bias weighed against the potential benefits of linkage (see Recommendations).

Research seeking consent to link survey with administrative data has encountered bias due to a significant portion of the sample responding negatively. Consent rates for this type of data linkage vary from 19% (McCarthy *et al.* 1999) to 97% (Rhoades & Fung 2004). These findings suggest linkage requests may be influenced by the mode of completion. We know respondents using a web survey mode are more likely to be regular and experienced users of the Internet, to have softer views regarding privacy and security, and therefore to have greater preferences for response in that mode (Al Baghal & Kelley 2016, de Leeuw 2005; Wenz *et al.* 2017). These Internet respondents are more likely to be social media users than CAPI or CATI respondents, which may increase their willingness to consent to data linkage. In contrast, some respondents are more comfortable in the presence of an interviewer. This is particularly the case with longitudinal surveys (Eisnecker & Kroh 2017). In longitudinal surveys, the effect of waves has also been found to lead to different consent outcomes (Sala *et al.* 2014). The differences in consent outcome could be even greater where mode of response changes across waves. How these factors impact on consent rates for linking social media with survey data have been explored only in a handful of studies.

The pilot research carried out by the ESRC Social Data Science Lab in partnership with NatCen and the University of Essex, examined the practicalities of linking Twitter user accounts to survey data. Twitter has similarities with other social media platforms, making these results applicable more broadly. The first practical obstacle to link Twitter user accounts (or other social media accounts) to survey data is respondent informed consent. If consent is granted, the respondent supplies their Twitter account information that acts as a unique identifier. If the account is not set to private, data can be passively collected historically and prospectively and linked to survey responses. If the account is set to private, the respondent must also consent to being followed (in the case of Twitter) by a survey representative before data can be accessed.

In our study we explored linking Twitter data to three surveys representative of the British adult population: the British Social Attitudes (BSA) survey (a cross-sectional survey that asked for consent to link Twitter data in 2015); the NatCen Panel; and the Understanding Society Innovation Panel (IP)

(the NatCen Panel and IP asked for consent to link Twitter data in 2017). By comparing these three surveys we were able to explore the impact that delivery mode had on consent outcomes, in addition to respondent demographics (Al Baghal *et al.* 2019).

The BSA survey (2015) was conducted entirely using CAPI. The achieved a sample size was 4,328, with a response rate of 51%. During interview, respondents were asked if they had a personal Twitter account. Those responding positively were asked for consent to link their Twitter data to their survey responses. If consent was granted, the respondent was asked for their Twitter account username (see Appendix for question wording).[2]

The NatCen Panel (July 2017) is probability-based and mixed-mode. It employs a sequential mixed-mode design, where members are first invited to participate online (using multiple points of contact by post, e-mail, and text). Non-completion after 2 weeks initiated contacted by telephone (where numbers were available) using CATI. The sample size was 2,184, with 82.2% completing on the web (1796) and 17.8% (388) completing via telephone. The survey response rate, i.e. the proportion of participants invited to take part completing the survey, was 60%. The question asking consent to link Twitter user account to survey data was based on the BSA question (see Appendix for question wording).

The IP Wave 10 (2017) is part the UK longitudinal household study, Understanding Society. It uses a multistage probability sample of persons and households. Response rates are calculated as completion rates among those responding at their initial wave of interview. At the initial wave IP1 (2008) the response rate was 52%. The IP4 (2011) refreshment sample response rate was 44%, and the initial IP7 (2014) response rate was 24%. The reinterview rates at IP10 for those interviewed at IP1 was 31%; for the IP4 refreshment sample the reinterview rate at IP10 was 48%; and the reinterview rate for the IP7 refreshment sample at IP10 was 62%. For IP10, those responding via the web could access the survey by via PC, tablet, or smartphone. 60% of web respondents completed via a PC, 29% via tablet, and 12% via smartphone. The IP10 refreshment sample was conducted only via CAPI. The consent request to link Twitter user accounts to survey responses was placed near the beginning of the survey. The consent question used was in the same form as the NatCen Panel (see Appendix for question wording).

In the NatCen panel and IP surveys the request to link Twitter data included the information required for informed consent: the reason for collecting the data; how the data will be used; the information that will be collected; and guarantees of security and privacy. Wording was broad ('*such as* your profile information…') to facilitate the generation of datasets for archiving and reuse, and information was kept simple to minimize cognitive burden and to avoid misunderstandings.

> *As social media plays an increasing role in society, we would like to know who uses Twitter, and how people use it. We are also interested in being able to add people's, and specifically your, answers to this survey to publicly available information from your Twitter account such as your profile information, tweets in the past and in future, and information about how you use your account.*
>
> *Your Twitter information will be treated as confidential and given the same protections as your interview data. Your Twitter username, and any information that would allow you to be identified, will not be published without your explicit permission.*

[2] Members of the BSA (2015) were invited to join the NatCen Panel. This resulted in some respondents being asked twice to consent to their Twitter user account being linked to survey data (albeit with different question wording). This allowed for comparisons between waves and survey mode (CAPI compared to telephone or web).

A series of hyperlinks were provided (to the respondent in web mode and to the interviewer in CATI mode) to more detailed information (full text in Appendix):

What information will you collect from my Twitter account?
What will the information be used for?
Who will be able to access the information?
What will you do to keep my information safe?
What if I change my mind?

We found that Twitter use in the UK is non-trivial. The surveys show Twitter usage in the UK adult population at 18% in 2015 (BSA) and 22% (IP) or 26% (NatCen Panel) in 2017. The BSA 51% response rate suggests that if all Twitter users consented to linkage, only 9.3% of the original sample would have linked data, presenting significant bias.

Consent	Total	Male	Female	<£1,800	>£1,800	Employed	Nonemployed	Higher Education Degree	Professional/ A Level	Other Education
Consented (%)	36.8	38.8	34.8	40.1	38.4	37.9	33.9	36.1	38.5	34.9
Base n	*791*	*400*	*391*	*211*	*326*	*570*	*221*	*288*	*291*	*206*

Table 3. BSA Twitter linkage consent rates: total and by respondent demographics

Table 3 provides a summary of data linkage consent rates for the BSA (2015) survey. The overall consent rate was 37% (291), with little variation by the demographics of respondents. However, age differences for consent did emerge. Younger respondents were significantly more likely to consent to Twitter data linkage than older respondents. While these results suggest any final linked data set may be biased due to the low consent rate, the homogeneity of response across demographics (but for age) suggests that non-consent bias by group may be minimal.

Overall, 26% of the NatCen Panel and 22% of IP respondents had Twitter accounts. Twitter usage varied significantly by response mode: in the NatCen Panel, 16% of telephone respondents and 28% of web respondents reported having a Twitter account. In the IP, 18% of CAPI respondents and 25% of web respondents indicated having a Twitter account.

Study	Overall	Interviewer Administered	Web	
NatCen panel	27.1% ($n = 151$)	34.4% ($n = 21$)	26.2% ($n = 130$)	($\chi^2_1 = 1.88$, $p = .170$)
Innovation panel	30.6% ($n = 131$)	40.5% ($n = 68$)	24.3% ($n = 63$)	($\chi^2_1 = 12.68$, $p < .0001$

Table 4. NatCen and IP Twitter linkage consent rates by mode

Table 4 provides a summary of data linkage consent rates for the NatCen Panel and the IP across mode of response. Web administered surveys yielded lower rates of consent compared to CAPI/CATI. The NatCen Panel and IP CAPI/CATI consent rates are similar to the BSA CAPI consent rate (37%). Based on results from all three surveys, Twitter users are disinclined to link their data to surveys.

Multivariate regression modelling estimated the impact of survey mode and respondent demographics on likelihood to consent to Twitter data linkage. In the NatCen Panel, holding all other factors constant, males and younger respondents were significantly more likely to consent to link their Twitter user accounts. In the IP, CAPI respondents were significantly more likely to consent than web respondents. It is likely that the physical presence of an interviewer increases chances for

consent, but as has been found for survey nonresponse, telephone acquiescence falls between face-to-face and web interviewing.

Linking consent rates and correlates for non-social media DTD data

DTD are also obtainable from non-social media sources, such as smartphone technology (e.g. GPS location and bio-readings) and apps. An Understanding Society Innovation Panel study found that consent to link smartphone DTD to surveys varied by hypothetical task. Of those who used smartphones in the sample, 28% said they would consent to install an app that anonymously tracked phone usage, 39% would consent to sharing GPS data, 61% would consent to sharing accelerometer data, and 65% would consent to taking and sharing photos or scan bar codes (Wenz *et al.* 2017). The Dutch Longitudinal Internet Studies for the Social Sciences Mobile Mobility study gained consent from 19% of panel members to passively collect geolocation via a time use survey app (Scherpenzeel 2017). Similarly, a Spanish study achieved a consent rate of 20% to share GPS data and 18% to install a website tracking app (Revilla *et al.* 2018).

A study in Germany asked those who owned smartphones in a nonprobability online panel about their willingness to consent to link their DTD to surveys via a series of vignettes (Keusch *et al.* 2019). Overall, 35% of respondents were willing to consent to share their DTD in all eight studies described in the vignettes. Consent varied by (i) duration of project, with shorter periods achieving higher consent rates; and (ii) monetary incentives, with the most generous offer (20 euro) increasing the consent rate from 20% (no reward) to 46%. The main reasons given for non-consent were privacy and security concerns (44%) followed by 'no incentive' and 'incentive too low' (17%), while the main reasons for consent were 'interest', 'curiosity' (39%) and an 'incentive' (26%). There was a greater willingness to take part in university sponsored linked DTD research (37%) compared to government sponsored research (33%), confirming findings from the UK (Williams *et al.* 2017). In another German study, respondents in the Labour Market and Social Security panel were asked to consent to link their DTD via a smartphone app (Kreuter *et al.* 2019). The app passively measured phone network quality and location; interaction history; social network characteristics; and activity data. The app also sent short surveys for respondents to complete. Overall, 16% of invited panel respondents consented to install the app. Of these, above 90% consented to at least one form of passive monitoring, with 71% consenting to all functions. Following GDPR, researchers provided the option for participants to revoke permissions at any point during the process. Within the app participants could check mark a passive monitoring process to activate and deactivate DTD sharing, but the vast majority did not change any settings following installation.

Ethical Issues

Beyond acquiring informed consent to link DTD to survey data, there remain a range of issues related to privacy and harm that must be addressed ahead of analysis, publication and storage. This section provides an overview of the ESRC Social Data Science Lab's most recent research that explores these issues as applied to the linking of social media, in particular Twitter data, to survey data.

A non-probability sample of Twitter users found that 80% of respondents expected to be asked for their consent ahead of their Twitter content being published, and over 90% stated they expected anonymity in publication (in particular female and black and minority ethnic tweeters, and those posting personal photographs) (Williams *et al.* 2017). Parents, females, and lesbian, gay and bisexual Twitter users, were more likely to expect to be asked for their informed consent. These patterns reflect those found in the Eurobarometer Attitudes on Data Protection Survey (2011) that showed three quarters of Europeans accepted that disclosing personal information was now a part of modern life, but only a quarter of respondents felt that they had complete control over their social media information, and 70% were concerned that their personal data may be used for a purpose other than for which they were archived. A clear majority of Europeans (75%) wanted to delete personal information on a website whenever they decide to do so, supporting the 'right to be forgotten' principle.

Unlike with smartphone DTD generated by the device's own systems or an app, social media data in their original form are publicly accessible, meaning individuals are readily identifiable. The ethical challenge of working with linked social media and survey data is how to maintain privacy without the removal of key information that may hinder analysis. If precautions are not taken, linking Twitter data to survey data is potentially problematic. Twitter user names, message text, and much of the metadata can result in respondent deanonymisation. Within our Lab we have considered how to provide access to linked Twitter data while ensuring voluntary participation (informed consent), minimising harm (disclosure control and security) and maximising value (archiving). A range of solutions have been identified, including secure remote access, secure on-site locations, and data controller only linkage. The latter solution ensures researchers do not gain access to original Twitter data and survey responses simultaneously. In relation to archiving, solutions must be found that align with social media company terms of use and GDPR, that include limited sharing and the removal of content deleted by users post collection.

Ensuring the conditions of consent are adhered to requires an understanding of the pratical and technical factors relating to the collection, analysis and storage of Twitter data. Over 150 data fields are associated with a single tweet. The collection of 100 tweets from one individual would result in over 15,000 cells of data, potentially including account name, person name, tweet text, number of followers, number followed, time zone, home village/town/city, location where tweets were posted, and device used to access Twitter. If data are enhanced via the use of predictive algorithms (as is possible with the ESRC funded COSMOS software), additional non-verified information can be generated, including gender, age and occupation (Sloan *et al.* 2013, 2015, Sloan 2017). Many of these attributes are unique to the tweet and the user, including those that may at first seem innocuous.

Table 5 shows the risk of identifying an individual by a sample of tweet attributes that are likely to be of interest to the social science researcher. The attributes with the highest risk of disclosure are easily identifiable (e.g. tweet text, tweet IDs, screen name and user IDs). There are several attributes that for many users are not high risk, but may become so for some under certain circumstances. For example, the time and date of the creation of an account is precise, meaning it can identify a user if it is unique (i.e. in the case where no other user created an account at that exact time). Profile descriptions can be altered by users, but are likely to remain unique and hence linkable to historic records of tweets. URLs in user profiles often refer to organisational or personal webpages that may identify an individual. Responses to posts may relate to partners, children, or work colleagues. Number of followers, followees, and tweets can lead to the identification of individual accounts where values are extreme.

While attributes with unique values are most likely to lead to disclosure, cross-referencing non-unique attributes can also increase the likelihood of individual identification. As with survey data, the chance of disclosure increases through the crosstabulation of several non-unique variables that result in low cell counts. For example, knowing a user has 1,345 followers is not likely to disclose them, but crosstabulating this attribute with the number the user follows, the number of posts they have liked and the number of lists to which they subscribe, reduces the cell count low figures. Adding the time stamp of the tweet to eradicate any variability introduced by the gap between the date of capture and the date of analysis, reduces that cell count even further, potentially to low single figures. A suitably motivated person could therefore use cross-referenced metric data with the time and date for when it was correct to derive a small group of users and in some cases pinpoint an individual. This example uses a limited number of Twitter attributes that are commonly known to users of the site. The amount and variety of data obtained through the Twitter Application Programming Interface (API) that would be used in a linked data study is significant. These data will be stored somewhere and hence accessible and usable in such a process if not secured using established standards.

Table 6 summarises four areas where data security should be considered when processing linked Twitter and survey data: systematic processing; data reduction; controlled access; and data deletion. Dependent on the nature of the social media data, these principles may apply to platforms other than Twitter. Figure 1 delineates one way of securely processing linked data, derived from the ADRN 'systematic processing' system. Initially, the data collected from the survey will include a unique ID,

the survey data, and the consenting panel member's Twitter account name (1). The first stage of processing splits these data – separating the identifying Twitter account name from the survey data into two datasets – (2) and (3) – with both carrying a unique ID which allows for re-linking. The next stage (4) involves accessing Twitter account names to request panel members' data from the Twitter API (4). Caution should be exercised at this point as the Twitter data are identifiable. Consideration should be given to where these data are stored and who has access at this stage. Preferably, a data reduction process, for example dropping Twitter account names and removing attributes not required in the analysis (5), is performed to minimise risk of disclosure. Once the required Twitter attributes have been downloaded (and the account details removed), the two datasets may be linked back together for analysis (6). This analysis should be conduct in a controlled access environment.

Table 5: Sample of tweet metadata and the risk of deanonymisation (adapted from Sloan *et al.* 2019)

Risk of identifying an individual:	Relating to:	Attribute:	Description:	Nature of risk:
HIGH	Tweet	Text	The actual text of the tweet	Unique content and user directly identifiable
HIGH	Tweet	id_str	The numeric (string) version of the unique identifier for this tweet	Unique content, directly identifiable - often deposited to allow other researchers to 'rehydrate' Twitter datasets
HIGH	User	screen_name	The screen name (aka handle) of a user	Screen name can change (dynamic) but is always unique, an individual identifier
HIGH	User	user_url	A URL given by the user, normally a link to a personal/organisational website	Not necessarily unique, but will be in some cases, not unusual for users to direct to personal websites
HIGH	User	Name	The self-defined name of the user	Not necessarily the name of a person, but often is
HIGH	User	user_id_str	The numeric (string) version of the unique identifier for this user	Unique identifier, directly identifies the user
HIGH	User	description	User-defined description of their account, often used as a 'bio'	Regardless of what the user writes, this is likely to unique to the individual
HIGH	User	user_created_at	Creation date and time of the user account to the second (in UTC) e.g. Tue Nov 23 12:46:54 +0000 2018	Potentially unique to the individual due to high level of temporal granularity, note that offset ('+0000') can be used to determine time zone (but see later comment on GDPR)
MEDIUM*	Geo	Lat	Latitude of tweet location	Precise latitude of where user was when they tweeted, potentially could be at home or work, alternatively may be commuting. Has considerable potential to locate individuals in low level geographies, but this is significantly reduced without longitude value. *risk is considerably higher with corresponding longitude
MEDIUM*	Geo	Lon	Longitude of tweet location	Precise longitude of where user was when they tweeted, potentially could be at home or work, alternatively may be commuting. Has considerable potential to locate individuals in low level geographies, but this is significantly reduced without latitude value. *risk is considerably higher with corresponding latitude
VARIABLE	Tweet	in_reply_to_screen_name	If the tweet is a reply to another tweet, this is the name of the original tweet's author	Evidence of Twitter correspondence with another unique user, may or may not represent someone in their network, often used for responding to public individuals (e.g. politicians) but could also be used to respond to users who are closely connected
VARIABLE	Tweet	in_reply_to_status_id_str	If the tweet is a reply to another tweet, this is the ID of the original tweet	Represents part of a conversation that the user is partaking in, could be used to identify an individual if number of responses to original tweet are small
VARIABLE	Tweet	in_reply_to_user_id_str	If the tweet is a reply to another tweet, this is the ID of the original tweet's author	Evidence of Twitter correspondence with another unique user, may or may not represent someone in their network, often used for responding to public individuals (e.g. politicians) but could also be used to respond to users who are closely connected

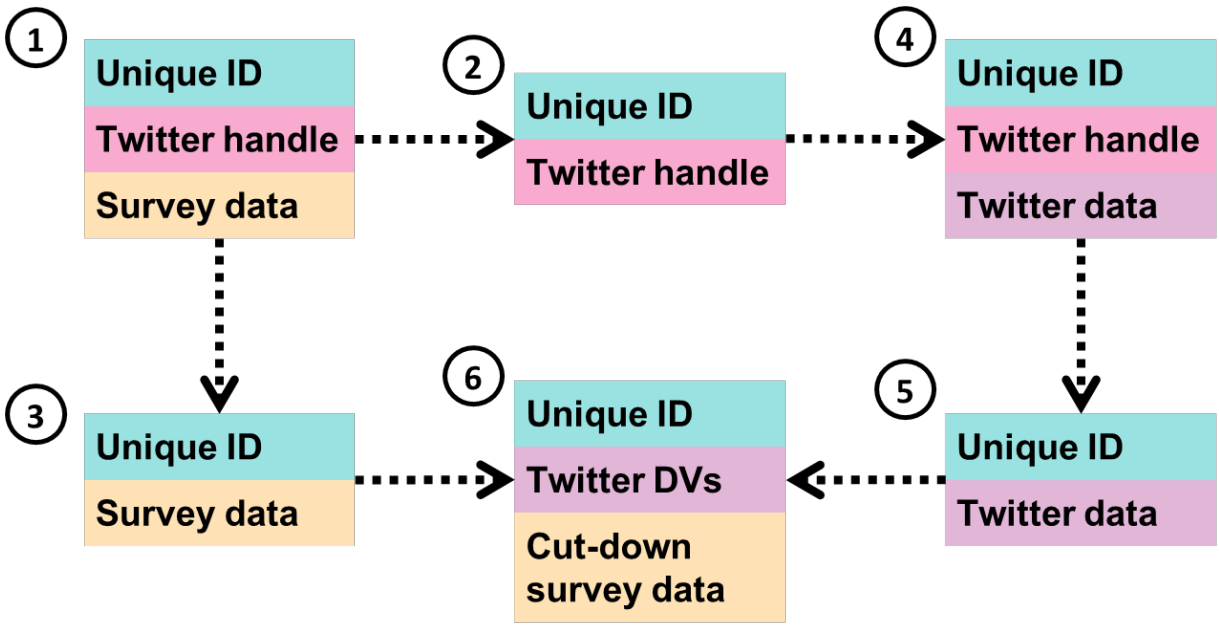
VARIABLE	Tweet	url	Wrapped URL corresponding to the value directly embedded into the raw tweet text	Depends where the URL points to, often to generic content (e.g. BBC News story) but could be to personal website or blog
VARIABLE	User	Location	The location defined by the user	May or may not represent where the user lives or works, but potentially could place user in a low level spatial unit
VARIABLE	Geo	full_name	Full name (string) of place e.g. 'San Francisco, CA'	Could lead to low level-spatial data if point coordinates, or user selection, results in identifying a city or town
VARIABLE	Geo	place_name	Short name (string) of place e.g. 'San Francisco'	Could lead to low level-spatial data if point coordinates, or user selection, results in identifying a city or town
LOW*	Tweet	retweet_count	The number of times a tweet has been retweeted	Changeable and dynamic, unlikely to be unique * unless extreme
LOW	Tweet	created_at	Creation date and time of the tweet to the second (in UTC) e.g. Tue Nov 23 12:46:54 +0000 2018	On average there are 6,000 tweets created every second, and difficult (if at all possible) to acquire all historic tweets made in a given second without access to the Twitter's enterprise product (100% feed). Note that offset ('+0000') could be used to determine time zone (but see later comment on GDPR)
LOW*	Tweet	favorite_count	The approximate number of times a tweet has been liked by other users	Changeable and dynamic, unlikely to be unique * unless extreme
LOW*	User	statuses_count	The number of tweets and retweets posted by the user	Changeable and dynamic, unlikely to be unique * unless extreme
LOW*	User	followers_count	The number of followers the user account currently has	Changeable and dynamic, unlikely to be unique * unless extreme
NEGLIGIBLE	Tweet	retweeted	Indicates whether the tweet has been retweeted by the user	Binary categorical variable, common practice to retweet
NEGLIGIBLE	Tweet	Source	The utility used to post the tweet (e.g. Tweets posted from the Twitter website have a source of 'web')	Unlikely to pose a risk as alternative Twitter posting tools are in widespread use
NEGLIGIBLE*	Tweet	Lang	The language of the tweet text (machine-detected)	Machine detection will allocate to one language or mark as 'undetected', will only identify a single language, might well not be the same as language of interface, can change with every tweet (dynamic) * but might result in 'low cell count problem' for minority languages
NEGLIGIBLE*	User	friends_count	The number of accounts this user is following	Changeable and dynamic, unlikely to be unique * unless extreme
NEGLIGIBLE	Geo	Country	Name of the country a tweet was issued from or is about	May be derived from an exact point coordinate (lat/long), or form a place selected by a user such as a city. In the latter, this may be the country of the place from where the user is tweeting from, or a place that they are tweeting about. Either way, on its own this represents a high-level geography
N/A	User	time_zone	The time zone of the user	If present will place the user in a large-scale geography, but in EU has been returned as 'null' (private field) due to GDPR

Note: * denotes an attribute that is, for the most part, likely to be non-disclosive except in rare cases

Table 6: Principles for security in linked Twitter and survey data (adapted from Sloan *et al.* 2019)

1. Systematic processing	Data should be managed in a systematic and considered manner. Based on the processes used for linking survey and administrative records (ADRN 2018), once initial consent has been collected, Twitter and survey data should be stored and processed separately until data linkage is required for analysis, to minimise the risk of disclosure.
2. Data reduction	Only the survey and Twitter data necessary for analysis should be made available for linkage as it is likely that not all data are required to address the study objectives. Reducing the linked variables may mitigate disclosure. Removing 'high-risk' variables will significantly reduce risk. In some cases derived variables may suffice for analysis. For example, while the analysis may require raw Tweet content initially, the linked analysis may only require a derived variable indicating whether or not a Tweet contained a reference to a particular topic, which is less likely to be individually identifiable.
3. Controlled access	Access to identifiable data should be limited to those who need it. Linked data should be held securely, and those with access should be documented and be appropriately trained.
4. Data deletion	Data should only be held for as long as is necessary for analysis to be conducted. Once the project is complete, as with other forms of personal data, data should be securely deleted or archived in line with the social media platform's terms of use.

Figure 1: Data Flow Diagram for Linking Survey and Twitter Data (Sloan *et al.* 2019)



Following completion of analysis/end of the study, linked data should be deleted or archived for re-use following the process outlined by Kinder-Kurlanda (2017). Twitter terms of use allow for the sharing and archiving of tweet and user IDs that can be 'rehydrated' via the API at a future point in time. This process ensures that any deletions by Twitter or users since the time of the original collection will be reflected in the 'rehydrated' dataset. From an ethical perspective this is positive, as we might view such deletions as a withdrawal of consent, and these cases should be excluded from the dataset. However, from a replication perspective, such deletions introduce missing data that is unlikely to be at random.

Even with full knowledge of the vast array of data provided by the Twitter API, changes in legal requirements, such as the recent introduction of GDPR, can alter what is and is not acceptable. Even if the regulatory context is relatively stable, social media platforms can (and do) change the nature of the data provided through APIs, the terms of use for users and the conditions listed in developer agreements.

Recommendations

- The is a need for empirical work on the benefits of linked DTD and survey data. For example, it remains largely unexplored how linked Twitter data might be used practically to mitigate survey non-response and to generate measures that replicate and possibly substitute survey items.

- Initial nonresponse and panel attrition, combined with the small percentage of Twitter users in the UK, mean low consent rates for linked Twitter and survey data studies will likely result in small numbers and bias. This issue may be partly mitigated in other DTD linkage studies given the increasing popularity of smartphones in the UK. Despite these limitations, DTD and survey linkage is likely worthwhile in some use cases. Further research should be funded that develops our understanding how DTD linkage consent rates differ by survey mode, wave, respondent demographics and other relevant factors.

- GDPR and research ethics requirements put a strong emphasis on dynamic informed consent. Being informed requires active engagement and continuous effort on the side of participants in DTD studies that use passive collection. Further research is needed on understanding how consent requests relating to DTD increases the cognitive burden on research respondents, and how this impacts on their ability to provide informed consent.

- Changes in legal (e.g. GDPR) and technical requirements (e.g. Twitter and Facebook rules on data access) can rapidly shift the DTD research landscape, increasing the risk that researchers make the wrong and/or risky decisions. Consideration should be given to the funding of a network (or the expansion of an existing network) that establishes and maintains an infrastructure (including training) that allows social researchers to use linked social media data and other DTD in a safe setting, with the proper security measures in place.

References

AAPOR. (2014) Social Media in Public Opinion Research: Report of the American Association for Public Opinion Research (AAPOR) Task Force on Emerging Technologies in Public Opinion Research available at: https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf

Al Baghal, T., & Kelley, J. (2016). The stability of mode preference: Implications for longitudinal survey design. *Methods, Data, Analyses*, 10, 143–166.

Al Baghal, T., Sloan, L., Jessop, C., Williams, M. L., & Burnap, P. (2019). Linking Twitter and survey data: The impact of survey mode and demographics on consent rates across three UK studies. *Social Science Computer Review*. doi:10.1177/0894439319828011

Beauchamp, N. (2017). Predicting and interpolating state-level polls using Twitter textual data. *American Journal of Political Science*, 61, 490–503. doi:10.1111/ajps.12274

Boase, J., & Ling, R. (2013). Measuring mobile phone use: Self-report versus log data. Journal of *Computer-Mediated Communication*, 18, 508–519. doi: 10.1111/jcc4.12021

Burnap, P., Gibson, R., Sloan, L., Southern, R. and Williams, M.L. (2016). 140 characters to victory?: Using Twitter to predict the UK 2015 General Election. *Electoral Studies.* 41, pp. 230-233.10.1016/j.electstud.2015.11.017

de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21, 233–255.

DiGrazia, J., McKelvey, K., Bollen, J., and Rojas, F. (2013) More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior. *PLoS One*, 8(11).

Edwards, A.M, Housley, W., Williams, M.L., Sloan, L., & Williams, M. (2013). Digital social research, social media and the sociological imagination: Surrogacy, augmentation and re-orientation. *International Journal of Social Research Methodology*, 16, 245-260.

Eisnecker, P. S., & Kroh, M. (2017). The informed consent to record linkage in panel studies: Optimal starting wave, consent refusals, and subsequent panel attrition. *Public Opinion Quarterly*, 81, 131–143.

Génois, M., Zens, M., Lechner, C., Rammstedt, B., & Strohmaier, M. (2019). Building connections: How scientists meet each other during a conference. arXiv preprint: 1901.01182

Ginsberg, J., Mohebbi, M., Patel, R., Brammer, L., Smolinski M et al. (2009) Detecting influenza epidemics using search engine query data. *Nature,* 457 1012-1014. DOI:10.1038/nature07634

Gorman, E., Leyland, A. H., McCartney, G., White, I. R., Katikireddi, S. V., Rutherford, L. and Gray, L. (2014). Assessing the representativeness of population-sampled health surveys through linkage to administrative data on alcohol-related outcomes. *American Journal of Epidemiology*, 180, 941–948.

Gray, L., White, I. R., McCartney, G., Katikireddi, S. V., Rutherford, L., Gorman, E., & Leyland, A. H. (2013). Use of record-linkage to handle non-response and improve alcohol consumption estimates in health survey data: A study protocol. *BMJ Open*, 3, e002647.

Guess, A. M. (2015). Measure for measure: An experimental test of online political media exposure. *Political Analysis*, 23, 59–75. doi:10.1093/pan/mpu010

Haenschen, K. (2019). Self-reported versus digitally recorded: Measuring political activity on Facebook. *Social Science Computer Review*, doi:10.1177/0894439318813586

Hofstra, B., Corten, R., van Tubergen, F., & Ellison, N. B. (2017). Sources of segregation in social networks: A novel approach using Facebook. *American Sociological Review*, 82, 625–656. doi:10.1177/ 0003122417705656

Hopp, T., Vargo, C. J., Dixon, L., & Thain, N. (2018). Correlating self-report and trace data measures of incivility: A proof of concept. *Social Science Computer Review*, doi:10.1177/0894439318814241

Jäckle, A., Lynn, P., & Burton, J. (2015). Going online with a face-to-face household panel: Effects of a mixed-mode design on costs, participation rates and data quality. *Survey Research Methods*, 9, 57–70.

Karlsen, R., & Enjolras, B. (2016). Styles of social media campaigning and influence in a hybrid political communication system: Linking candidate survey data with Twitter data. *The International Journal of Press/Politics*, 21, 338–357. doi:10.1177/1940161216645335

Keusch, F., Antoun, C., Couper, M. P., Kreuter, F., & Struminskaya, S. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly*. doi.org/10.1093/poq/nfz007

Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J. and Morstatter, F. (2017). Archiving Information from Geotagged Tweets to Promote Reproducibility and Comparability in Social Media Research. *Big Data & Society*, 1-14.

Kreuter, F., Haas, G. -C., Keusch, F., Ba¨hr, S., & Trappmann, M. (2019). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*, doi:10.1177/0894439318816389

McCarthy, D. B., Shatin, D., Drinkard, C.R., Kleinman, J. H., & Gardner, J. S. (1999). Medical records and privacy: Empirical effects of legislation. Health Services Research, 34, 417–25.

Mellon, J. (2014). Internet search data and issue salience: The properties of Google Trends as a measure of issue salience. Journal of Elections, *Public Opinion and Parties*, 24, 45–72. doi:10.1080/17457289.2013.846346

Moat, H., Preis, T., Olivola, C., Liu, C., Chater, N. (2014) Using big data to predict collective behaviour in the real world. *Behavioral and Brain Sciences*, 37:1 92–93

Möller, J., van de Velde, R. N., Merten, L., & Puschmann, C. (2019). Explaining online news engagement based on browsing behavior: Creatures of habit? *Social Science Computer Review*, doi:10.1177/0894439319828012

Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018). Networks of audience overlap in the consumption of digital news. *Journal of Communication*, 68, 26–50. doi:10.1093/joc/jqx007

Murphy, J., Landwehr, J., & Richards, A. (2013). Using Twitter to Predict Survey Responses. Paper presented at the *Midwest Association of Public Opinion Research conference*.

Nelson, J. L., & Webster, J. G. (2017). The myth of partisan selective exposure: A portrait of the online political news audience. *Social Media + Society*, 3. doi:10.1177/2056305117729314

Noguchi, T., Stewart, N., Olivola, C., Moat, H. and Preis, T. (2014) Characteristing the Time-Perspective of Nations with Search Engine Query Data. *PLoS One,* 9:4 e95209. DOI:10.1371/journal.pone.0095209

O'Connor, B., Balasubramanyan, R., Smith, N. A., & Routledge, B. R. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (pp. 122–129). Palo Alto, CA: AAAI Press.

Pasek, J., McClain, C. A., Newport, F., & Marken, S. (2019). Who's tweeting about the president? What big survey data can tell us about digital traces. *Social Science Computer Review*, doi:10.1177/0894439318822007

Quinlan, S., Gummer, T., Roßmann, J., & Wolf, C. (2018). "Show me the money and the party!"—Variation in Facebook and Twitter adoption by politicians. *Information, Communication & Society*, 21, 1031–1049. doi: 10.1080/1369118X.2017.1301521

Revilla, M., Couper, M. P., & Ochoa, C. (2018). Willingness of online panelists to perform additional tasks. *Methods, Data, Analyses*, doi:10.12758/mda.2018.01

Rhoades, A. E., & Fung, K. (2004). Self-reported use of mental health services versus administrative records: Care to recall? *International Journal of Methods in Psychiatric Research*, 13, 165–75.

Sala, E., Knies, G., & Burton, J. (2014). Propensity to consent to data linkage: Experimental evidence on the role of three survey design features in a UK longitudinal panel. *International Journal of Social Research Methodology*, 17, 455–473.

Scherpenzeel, A. (2017). Mixing online panel data collection with innovative methods. In S. Eifler & F. Faulbaum (Eds.), *Methodische Probleme von Mixed-Mode-Ansa¨tzen in der Umfrageforschung* (pp. 27–49). Wiesbaden, Germany: Springer

Seresinhe, C. I., Preis, T., MacKerron, G. and Moat, H. S. (2019) Happiness is greater in more scenic locations, *Scientific Reports*, 9, 1, 4498

Sloan, L. (2017) Who tweets in the United Kingdom? Profiling the Twitter population using the British social attitudes survey 2015. *Social Media + Society* 3(1), pp. 1-11.

Sloan, L., Jessop, C., Al Baghal, T. and Williams, M.L. (2019) Linking Survey and Twitter Data: Informed Consent, Disclosure, Security and Archiving. *Journal of Empirical Research on Human Research Ethics*

Sloan, L.et al. (2013) Knowing the Tweeters: Deriving sociologically relevant demographics from Twitter. *Sociological Research Online*, 18(3), article number: 7.

Sloan, L.et al. (2015) Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *Plos One*, 10(3), article number: e0115545.
Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018). Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter. *Political Communication*, 35, 50–74. doi:10.1080/10584609.2017.1334728
Sugie, N. F. (2018). Utilizing smartphones to study disadvantaged and hard-to-reach groups. *Sociological Methods and Research*, 47, 458–491.
Thorson, K., Medeiros, M., Cotter, K., & Pak, C. (2018). Advertising categories as clues about political content exposure on Facebook. Paper presented to the *American Political Science Association Conference*, Boston, MA.
Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., & Tucker, J. A. (2015). Political expression and action on social media: Exploring the relationship between lower- and higher-threshold political activities among Twitter users in Italy. *Journal of Computer-Mediated Communication*, 20, 221–239. doi:10.1111/jcc4.12108
Vraga, E. K., & Tully, M. (2018). Who is exposed to news? It depends on how you measure: Examining self- reported versus behavioral news exposure measures. *Social Science Computer Review*, doi:10.1177/ 0894439318812050
Wells, C., & Thorson, K. (2015). Combining big data and survey techniques to model effects of political content flows in Facebook. *Social Science Computer Review*, 35, 33–52. doi:10.1177/0894439315609528
Wenz, A., Jäckle, A., & Couper, M. P. (2017). Willingness to use mobile technologies for data collection in a probability household panel. *Understanding Society Working Paper*, 2017-10.
Williams, M. L. and Burnap, P. (2016) Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology,* 56(2):211-238
Williams, M. L., Burnap, P. and Sloan, L. (2017) Crime Sensing with Big Data: The Affordances and Limitations of using Open Source Communications to Estimate Crime Patterns. *British Journal of Criminology,* 57(2):320-340.
Williams, M.L., Burnap, P., Javed, A., Liu, H. and Ozalp, S. (2019) Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *British Journal of Criminology*.

Appendix

Survey question wordings for consent to link Twitter data to survey data (BSA, NatCen Panel and IP)

BSA (2015)

Do you have a personal Twitter account?
Yes
No

IF Yes
We are interested in being able to link people's answers to this survey to the ways in which they use Twitter. We would also like to know who uses Twitter. A research project about who and how people use Twitter is being conducted by a team of researchers at Cardiff University. Are you willing to tell me the name of your personal Twitter account and for this to be passed to researchers at Cardiff University, along with your answers to this survey? Your Twitter name would not be published.
Yes
No

IF Yes
INTERVIEWER: Please enter the respondent's Twitter name here
Open Question (Maximum of 100 characters)

NatCen Panel (July 2017)

Do you have a personal Twitter account?
Yes
No

IF Yes
As social media plays an increasing role in society, we would like to know who uses Twitter, and how people use it. We are also interested in being able to add people's, and specifically, your answers to this survey to publicly available information from your Twitter account such as your profile information, tweets in the past and in future, and information about how you use your account. Your Twitter information will be treated as confidential and given the same protections as your interview data. Your Twitter username, and any information that would allow you to be identified, will not be published without your explicit permission.

Are you willing to tell me your personal Twitter username and for your Twitter information to be added to your answers to this survey?
Yes
No

HELP SCREENS AVAILABLE
HELP SCREEN: <u>What information will you collect from my Twitter account?</u>
We will only collect information from your Twitter account that is publicly available. This will include information from your account (such as your profile description, who you follow, and who follows you), the content of your tweets (including text, images, videos and web links), and background information about your tweets (such as when you tweeted, what type of device you tweeted from, and the location the tweet was sent from).We will collect information from your past tweets (up to the last 3,000) and will update this with information from more recent tweets on a regular basis.

HELP SCREEN: <u>What will the information be used for?</u>

The information will be used for social research purposes only. Adding your Twitter information and your survey answers will allow researchers from universities, charities and government to better understand your experiences and opinions. For example, using extra information from your Twitter account, researchers can start to:

- Understand who uses Twitter and how they use it
- See what Twitter information can tell us about people, and how accurate it is
- Know what people in the UK are saying about things we don't ask in our survey
- Look at additional information related to questions asked in the survey

HELP SCREEN: Who will be able to access the information?
Matched data which includes both your survey answers and Twitter information will be made available for social research purposes only. Researchers who want to use your matched Twitter and survey information must apply to access it and present a strong scientific case to ensure that the information is used responsibly and safely. Matched statistical information from your Twitter account which you cannot be identified from (e.g. how often you Tweet, or whether you follow any politicians) will have the same access controls as your other survey answers. At no point will any information that would allow you to be identified be made available to the public

HELP SCREEN: What will you do to keep my information safe?
All information we collect will be held in accordance with the Data Protection Act 1998. Because Twitter information is public data that anyone can search, it is impossible to anonymise completely. To keep your information safe, researchers will only be able to access the matched survey answers and detailed Twitter information in a secure environment set up to protect this type of data. Only approved researchers who have gone through special training may access this information, and they will have to apply to do so. Statistical information from your Twitter account which you cannot be identified from (e.g. how often you Tweet, or whether you follow any politicians) will have the same level of protection as your other survey answers.

HELP SCREEN: What if I change my mind?
This information will be collected and stored for as long as they are useful for research purposes, or until you contact us to withdraw your permission. You can do this at any time by emailing us at panel@natcen.ac.uk or calling 0800 652 4569, and do not have to give a reason.
{END OF HELP SCREENS}

IF Yes
What is your Twitter username?
SOFTCHECK: "Twitter usernames must begin with an @ character, followed a maximum of 15 characters (A-Z, a-z, 0-9, underscore), no word spaces. Please check and amend."

IP10 (2017)

Do you have a personal Twitter account?
Yes
No

IF Yes
We would like to know who uses Twitter, and how people use it. We are also interested in being able to add people's answers to this survey to publically available information from your Twitter account such as your profile information, tweet content, and information about how you use your account. Your Twitter information will be treated as confidential and given the same protections as your interview data. Your Twitter username, and any information that would allow you to be identified, will not be published without your explicit permission. Are you willing to tell me the name of your personal Twitter account and for your Twitter information to be linked with your answers to this survey?
Yes
No

HELP SCREENS AVAILABLE
HELP SCREEN: What information will you collect from my Twitter account?
We will only collect information from your Twitter account that is publically available. This will include information from your account (such as your profile description, who you follow, and who follows you), the content of your tweets (including text, images, videos and web links), and background information about your tweets (such as when you tweeted, what type of device you tweeted from, and the location the tweet was sent from). We will collect information from your past tweets (up to the last 3,000) and will update this with information from more recent tweets on a regular basis. This information will be collected and stored for as long as they are useful for research purposes, or until you contact us to withdraw your permission. You can do this at any time, and do not have to give a reason.

HELP SCREEN: What will the information be used for?
The information will be used for social research purposes only. Adding your Twitter information and your survey answers will allow researchers from universities, charities and government to better understand your experiences and opinions. For example, using extra information from your Twitter account, researchers can start to:
* Understand who uses Twitter and how they use it
* See what Twitter information can tell us about people, and how accurate it is
* Know what people in the UK are saying about things we don't ask in our survey
* Look at additional information related to questions asked in the survey

HELP SCREEN: Who will be able to access the information?
Researchers who want to use matched Twitter and survey information must apply to access it and present a strong scientific case to ensure that the information is used responsibly and safely. Matched statistical information from your Twitter account which you cannot be identified from (e.g. how often you Tweet, or whether you follow any politicians) will have the same access controls as your other survey answers.

HELP SCREEN: What will you do to keep my information safe?
Matched statistical information from your Twitter account which you cannot be identified from (e.g. how often you Tweet, or whether you follow any politicians) will have the same level of protection as your other survey answers.

IF Yes
What is your Twitter username (e.g. @usociety)?
Soft check: Twitter username does not being with '@' or contains spaces **"**Please check and amend. Twitter usernames should begin with an @ character and should not contain any spaces."

Ex Ante Linking	Ex Post Linking
(A) Aggregate level • Analysis of audience overlaps (e.g., Mukerjee et al., 2018; Nelson & Webster, 2017) • Analysis of aggregate audience statistics (e.g., political ideology, Nelson & Webster, 2017)	(B) Aggregate level Linking survey responses to digital trace data . . . • Temporally: both are generated during the same time period (e.g., Mellon, 2014; O'Connor et al., 2010; Stier et al., 2018) • Topically: both focus on the same topic (e.g., Pasek et al., 2019) • Geographically: both can be located within same geographic area (e.g., Beauchamp, 2017)
	(C) Public actors Link publicly available digital trace data of public actors (e.g., politicians or organizations) to their survey responses (e.g., Karlsen & Enjolras, 2016; Quinlan et al., 2017)
(D) Individual level Ask individuals in surveys for informed consent to record in real time: • Website visits (e.g., Guess, 2015; Jürgens et al., 2019; Möller et al., 2019; Vraga & Tully, 2018) • Smartphone data (e.g., Boase & Ling, 2013; Jürgens et al., 2019; Kreuter et al., 2019) • Sensor data (e.g., Génois, Zens, Lechner, Rammstedt, & Strohmaier, 2019)	(E) Individual level Ask individuals in surveys for informed consent to collect their historical digital trace data . . . • From social media APIs (e.g., Al Baghal et al., 2019; Haenschen, 2019; Hofstra, Corten, van Tubergen, & Ellison, 2017; Hopp, Vargo, Dixon, & Thain, 2018; Vaccari et al., 2015; Wells & Thorson, 2015) • via data donation, for example, personal Google or Facebook histories (e.g., Thorson et al., 2018)

Table 2: Linking types with examples from the literature

Source: Stier, S., Breuer, J., Siegers, P., and Thorson, K. (2019) 'Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field', *Social Science Computer Review*, DOI: 10.1177/0894439319843669