NatCen

Social Research that works for society

Linking Survey and Twitter Data

Ethics, consent, anonymity, archiving and sharing

July 2019

Contents

- Why link survey and Twitter data?

- Consent

- Anonymisation & data security

- Archiving & sharing

- Looking ahead

Acknowledgments

- Luke Sloan – University of Cardiff

- Tarek Al Baghal – University of Essex

- Matthew Williams – University of Cardiff

Background

- Lead for the 'New Social Media, New Social Science' network
- Research Director at NatCen Social Research

Why link Twitter & Survey data?

- Continued & increasing interest in using social media data for social research

  - A subject of interest and a new lens to view the world through
  - Large volumes of data produced in real-time in multiple media
  - 'Natural' data accessible for free

- But they continue to have their drawbacks

  - Representativeness of the sample
  - Data are not created *for* your research question
  - Ethical questions around lack of consent & anonymity
  - Costs of data processing and analysis

CARDIFF UNIVERSITY

University of Essex

NatCen Social Research

Why link Twitter & Survey data?

Linking survey & social media data attempts to offer survey data the benefits of social media data, while addressing some of their issues:

- Collecting informed consent
- Putting sample in context of the population
- Understanding whose data you are analysing
- Validating machine-based classifications

Should be viewed as a **complementary** methodology

Where do ethics come in?

Three particular areas of consideration:

- Consent to link the data

- Processes for linking the data

- Access to, archiving, and sharing the data

Asking for informed consent

- Need to ask consent to access & link Twitter data

  - We have the opportunity

  - Linking data creates additional risk

- Balance level of detail:

  - Enough detail that consent is <u>informed</u>

  - Not so much that people are overwhelmed & don't read/understand it

Consent question wording

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.

Consent question wording

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.

CARDIFF UNIVERSITY   University of Essex   NatCen Social Research

Consent question wording

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.

Consent question wording

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, <u>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.</u>

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.

CARDIFF UNIVERSITY

University of Essex

NatCen Social Research

Consent question wording

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.

CARDIFF UNIVERSITY

University of Essex

NatCen Social Research

Consent question wording

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.

CARDIFF UNIVERSITY

University of Essex

NatCen Social Research

Consent question help-links

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?

Tested in three studies in GB

- British Social Attitudes (BSA) 2015*

  - Cross-sectional sample

  - Face-to-face fieldwork

- NatCen Panel (NCP) July 2017

  - Longitudinal sample (recruited from BSA 2015 & 2016)

  - Sequential mixed-mode fieldwork (web/telephone)

- Understanding Society Innovation Panel (IP) wave 10

  - Longitudinal sample

  - Face-to-face and sequential mixed-mode fieldwork (web/face-to-face)

*a different question wording was used in this study

Consent rates
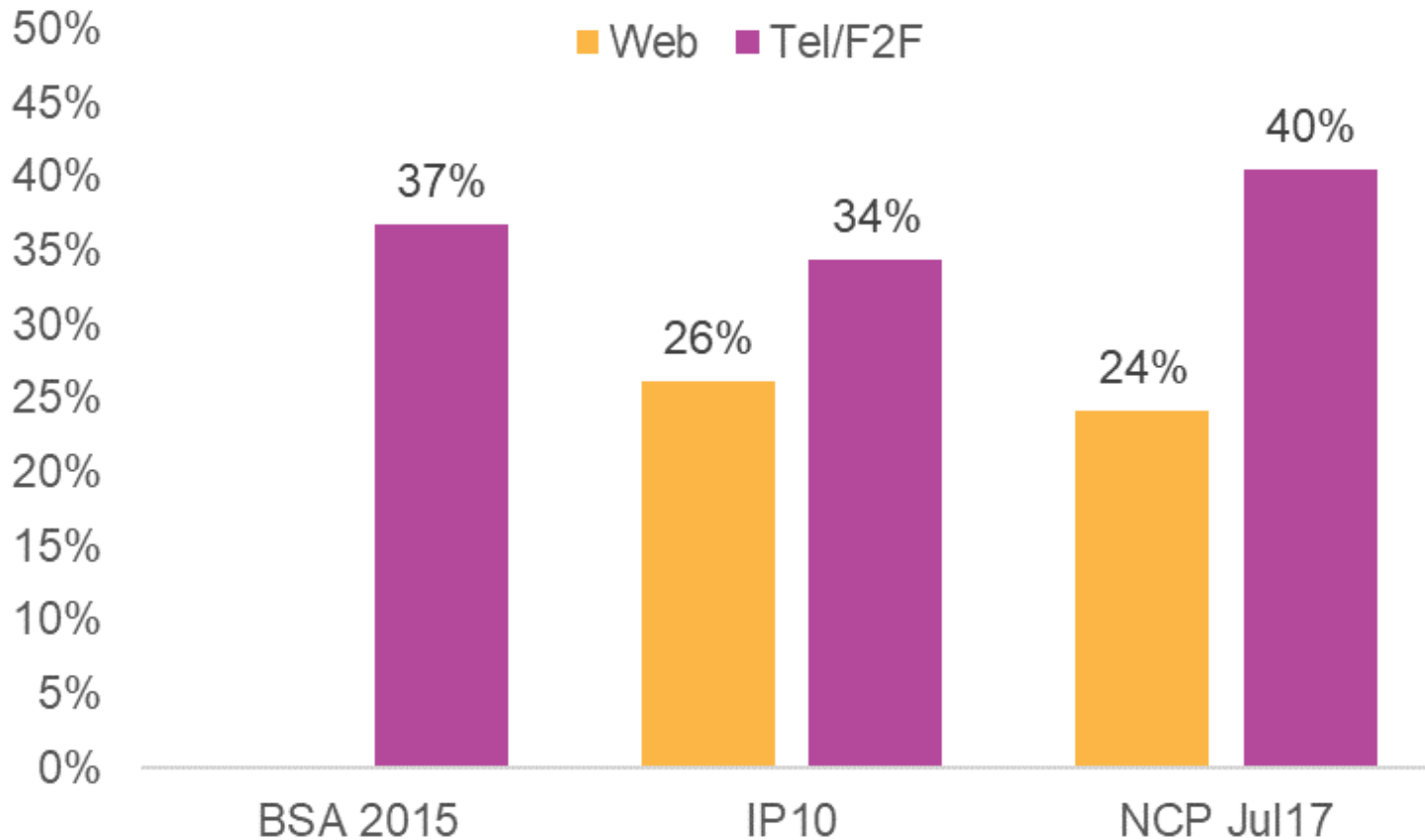
Base: Adult Twitter users in GB: BSA (F2F: 791);
IP10 (Web: 497; F2F: 61); NCP (Web: 260; Tel: 168)

Anonymisation?

- Surveys rely on anonymisation, but raw Twitter data are inherently identifiable, and it is in this raw format that they are useful

- Once the two are linked…

    - Survey data are no longer anonymous

    - Twitter data are no longer 'public'

- Need to step back from 'tool' of anonymisation, and think about other approaches to minimising risk of harm to participants

    - Systematic processing

    - Data reduction

    - Controlled access

    - Data deletion

Systematic processing
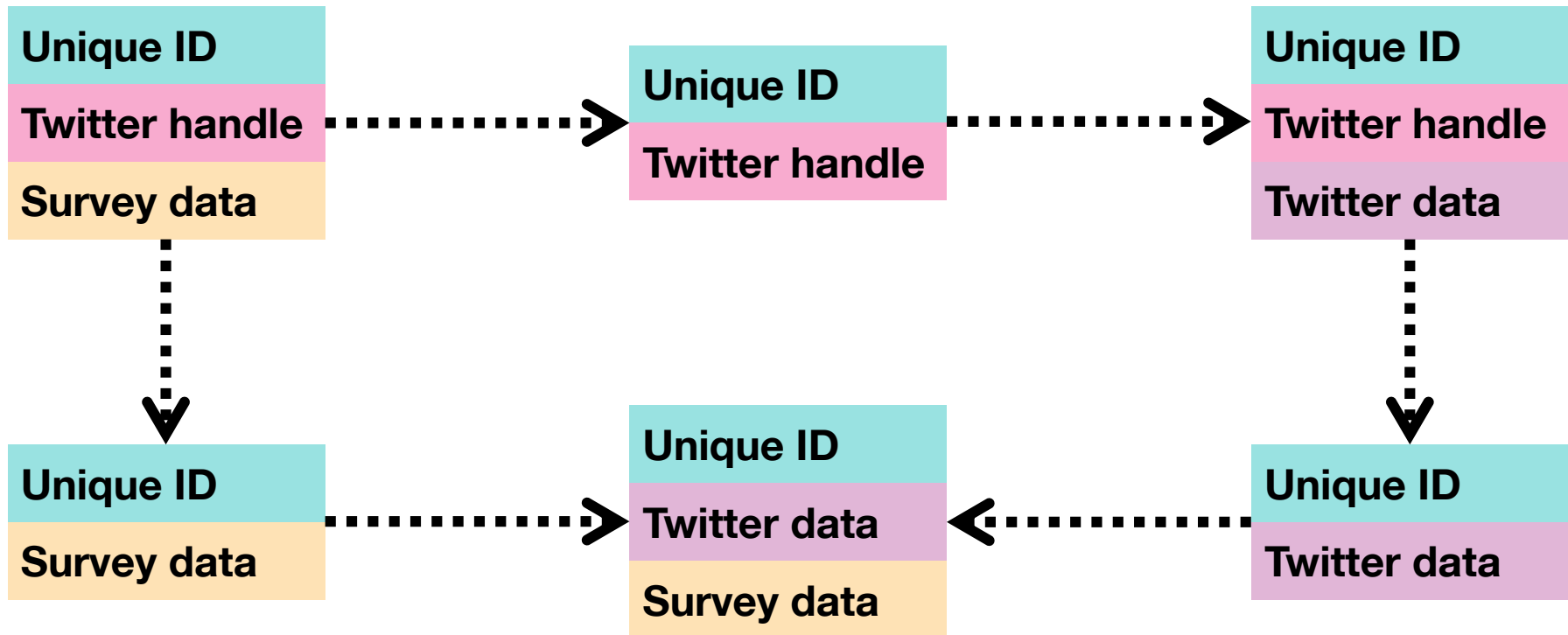
Data reduction
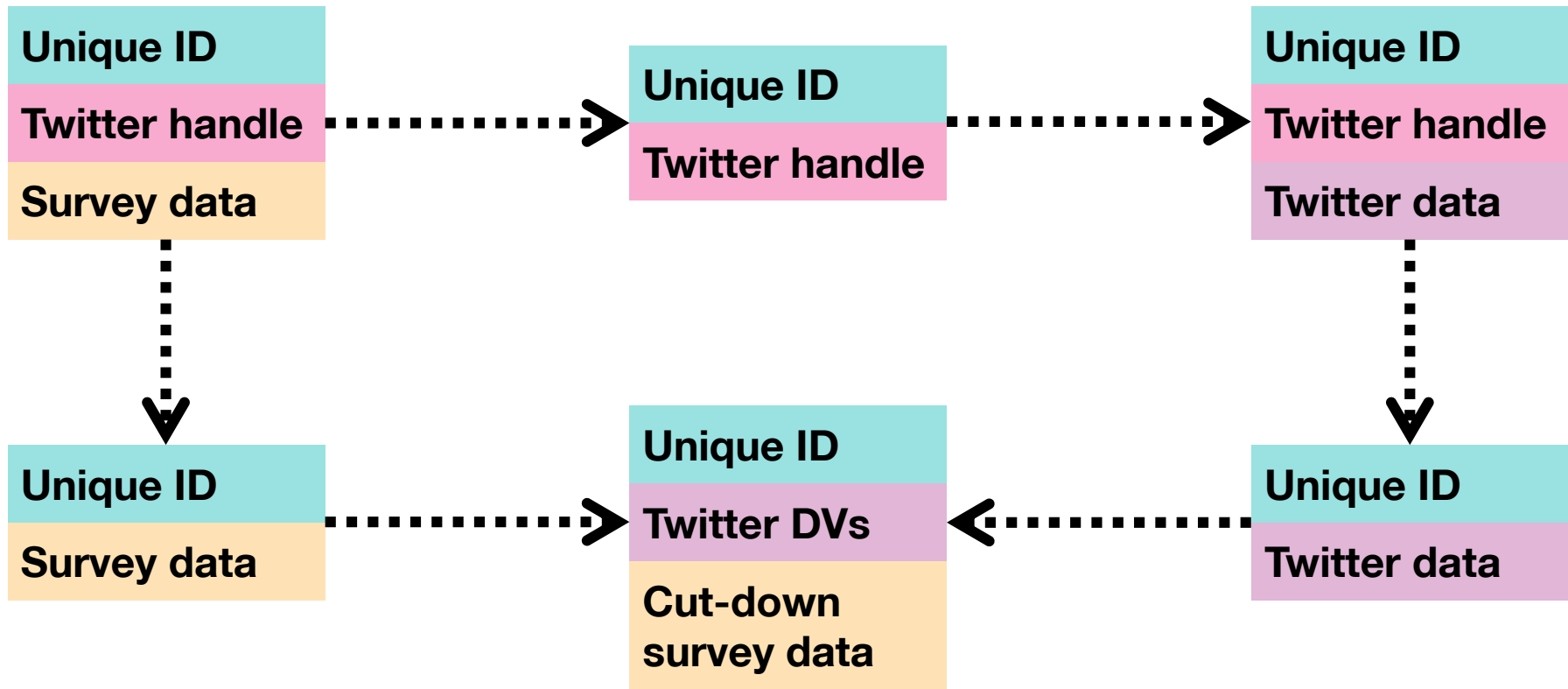
Controlled access

- Access should be justified & limited to those who need it

- Documentation of access

- Training & accreditation for researchers using the data

- Access in a secure environment
  - Offline (if possible)
  - Not able to take data away (without review)
- Data deletion

Archiving & sharing

- Archiving and sharing of data is important for the replication of results and to maximise value of data

- Similar issues (and solutions?) with initial data processing

- But additional hurdles:

    - Sharing datasets and Twitter's ToS

    - Deleted Tweets/withdrawn consent?

    - Who is responsible for maintaining the data?

Looking ahead

- Linking is feasible, but key challenge of low consent rates
    - Understand why & how to address
- Next step of demonstrating value in applied setting
    - Practical applications for understanding online & offline society
- Infrastructure & processes for archiving & sharing Twitter data
- Expansion to other platforms; is Twitter 'the future'?
    - The 'digital eco-system'
    - New & changing digital data forms

CARDIFF UNIVERSITY   University of Essex   NatCen Social Research

Full papers

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*.

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

Contact

Curtis Jessop

Research Director

curtis.jessop@natcen.ac.uk

@CurtisJessop

NatCen
Social Research that works for society