Ethics in Social Media Research

Matthew Williams

Social Data Science Lab
Cardiff University

@mattlwilliams
@socdatalab
• Interdisciplinary approach the challenges of “Big Social Data”

• Collaboration between Social and Computer Scientists at Cardiff University

• Focus on the methodological, technical, theoretical, and empirical dimensions of Big Social Data in research, policy and operational contexts

• Development of new methodological insight, tools and technical/data solutions for UK academia and public sector
We distribute and use the COSMOS Platform
2 affected attack bomb bombing bombs breaking crazy day do everyone explosion explosions finish go going good happened hope I'm injured line love luck near news now out
Lab Software.

- COSMOS Platform Social Data Transformation
  - Word Frequency
  - Point data frequency over time
  - Social Network Analysis
  - Geospatial Clustering
  - Sentiment Analysis
  - Demographic Analysis (gender, location, age, occupation/social class)
  - Data linking with any open data source (e.g. census, RSS)
Lab Data.

• Lab Data Collection and Curation
  – Persistent connection to Twitter 1% Stream (~4 billion)
  – Geocoded tweets from UK (~200 million annually)
  – Bespoke keyword-driven Twitter collections (crime & security, health & wellbeing, political opinion)
  – ONS & Police API
  – Drag and drop RSS
  – Import CSV/JSON
  – Web enabled so push/pull data from anywhere
Projects.

Previous and ongoing Social Data projects involving Lab staff:
Digital Social Research Tools, Tension Indicators and Safer Communities: (ESRC DSR) 2011-12

Supporting Empirical Social Scientific Research with a Virtual Research Environment (JISC) 2013

Small items of research equipment at Cardiff University (EPSRC) 2013

Hate Speech and Social Media: Understanding Users, Networks and Information Flows (ESRC Google) 2012-14

Social Media and Prediction: Crime Sensing, Data Integration and Statistical Modelling (ESRC NCRM) 2013-15

Understanding the Role of Social Media in the Aftermath of Youth Suicides (Department of Health) 2013-15

Scaling the Computational Analysis of “Big Social Data” & Massive Temporal Social Media Datasets (HPC Wales) 2015

Digital Wildfire: (Mis)information flows, propagation and responsible governance (ESRC Global Uncertainties) 2014-16

Public perceptions of the UK food system: public understanding and engagement, and the impact of crises and scares (ESRC/FSA) 2014-16
Social Media Research Ethics Survey (2013-15)

- Online survey
- N=564 and over 15,000 words
- Sections on:
  - Frequency of social media use (multiple platforms)
  - Type of use (type of posts: text, image, video; content of posts)
  - Awareness of Terms of Service
  - Attitudinal questions on informed consent and social media
  - Attitudinal questions on social media research and anonymisation
  - Attitudinal questions on social media research and type of researcher (commercial, government, police & university)
- Demographic questions
- Open questions to capture qualitative responses
Are you aware that social media networks have Terms of Service?

- Yes: 93.7%
- No: 5.4%

Have you ever read any ToS for social media networks to which you belong?

- Yes: 50.9%
- No: 47.9%

Did you know that when accepting ToS you consented to your posts to be accessed by third parties?

- Yes: 74.2%
- No: 24.1%
Concern – Commercial.

How concerned are you about your social media information being used by Commercial Organisations?

- Very concerned
- Quite concerned
- Slightly concerned
- Not at all concerned

How concerned are you about your social media information being used by Commercial Organisations?

- Net Experts
- Net Obsessives
- Single Tweeters
- LGB Tweeters
- Older Tweeters
- Female Tweeters
- Twitter Addicts
- High Income Tweeters
Concern – Government.

How concerned are you about your social media information being used by Government Organisations?

- Very concerned
- Quite concerned
- Slightly concerned
- Not at all concerned

X more likely to agree/disagree

How concerned are you about your social media information being used by Government Organisations?

- Net Experts
- Net Obsessives
- LGB Tweeters
- Older Tweeters
- Student Tweeters
- High Income Tweeters

X more likely to be concerned
Concern – Police.

How concerned are you about your social media information being used by Police?

- Very concerned
- Quite concerned
- Slightly concerned
- Not at all concerned

How concerned are you about your social media information being used by Police?

- Net Obsessives
- LGB Tweeters
- Older Tweeters
- Aware of T&S Consent
- Student Tweeters
Concern – Universities.

How concerned are you about your social media information being used by University Researchers?

- Very concerned
- Quite concerned
- Slightly concerned
- Not at all concerned

How concerned are you about your social media information being used by University Researchers?

- Net Obsessives
- Parent Tweeters
- Single Tweeters
- Female Tweeters
- Aware of T&S Consent
- Student Tweeters
Consent – Universities.

I don't mind if university researchers use my Tweets as data without consent

- Agree
- Tend to Agree
- Tend to disagree
- Disagree

X more likely to agree/disagree

I don't mind if university researchers use my Tweets as data without consent

- Aware of T&S Consent
- Net Experts
- Student Tweeters
- Single Tweeters
- BME Tweeters
- LGB Tweeters
- Female Tweeters

X more likely to agree/disagree
Anonymity – Universities.
Maintain the Integrity of Twitter’s Products

- @username must always be displayed (and name if possible) with tweet text
- Respond to content changes such as deletions or public/private status of tweets
- Do not modify, translate or delete a portion of the Content

Respect Users’ Privacy and get the user’s express consent before you do any of the following:

- Take any actions on a user’s behalf, including posting Content and modifying profile information
- Store non-public Content such as direct messages or other private or confidential information
- Share or publish protected Content, private or confidential information
Guidelines for Displaying Tweets in Broadcasts (exhibition, distribution, transmission, reproduction, public performance or public display of Tweets by any and all means of media):

- Include the user’s name and Twitter handle (@username) with each Tweet
- Use the full text of the Tweet. You may edit or revise Tweet text only as necessary due to technical or medium limitations (e.g., removing hyperlinks)
- Don’t delete, obscure, or alter the identification of the user. You may show Tweets in anonymous form in exceptional cases such as concerns over user privacy
- In some cases, permission from the content creator may still be necessary, as Twitter users retain rights to the content they post

Twitter Best Practices for Media (static uses and publication):

- Show name, @username, unmodified Tweet text and the Twitter bird nearby, as well as a timestamp
- If displaying Tweets, make sure they are real, from legitimate accounts and that you have permission from the author when necessary
- Display the associated Tweet and attribution with images or media
- If showing screenshots, only show your own profile page, the @twitter page, the Twitter “About” page or a page you have permission from the author to show
We follow the ESRC’s Framework for Research Ethics
All projects undergo Research Committee Review
Any significant changes to research design following are reported back to the Committee for re-approval
We abide by Twitter’s Developer Policy and Developer Agreement
In most cases we only publish in research outputs aggregate information based on data derived from the Twitter APIs
In research outputs we never directly quote identifiable individual Twitter users without informed consent. Where informed consent cannot be obtained we represent the content of tweets in aggregate form (e.g. topic clustering, wordclouds) and themes (decontextualised examples and descriptions of the meaning or tone of tweet content). These forms of representation preclude the identification of individual Twitter users, preserving anonymity and confidentiality
In research outputs we do directly quote from identifiable Twitter accounts maintained by public and commercial organisations (e.g. government departments, law enforcement, local authorities, companies) without seeking prior informed consent
We never share data gathered from Twitter APIs for our research outside of the Lab or project partners
Lab Risk Assessment.

**Low risk** – Tweet is from official/institutional account: Publish without seeking consent in most cases.

**Medium risk** – Tweets are from individual users and contain mundane information of a non-sensitive nature: Must contact the user (direct message/@mention/email) informing them of the intent to publish; unless the user opts out consider as permission to publish.

**High risk** – Tweets are from individual users and contain sensitive information (overly personal, abusive etc.). Must contact the user (direct message/@mention/email) and ask their permission to publish. Only publish if consent is received.

**High risk** – Tweet has been deleted precluding publication under Twitter Developer Agreement/Policy.

**High Risk**– Tweet is from a deleted account meaning it has been deleted precluding publication under Twitter Developer Agreement/Policy.