# Legally and Ethically Using Social Media Data in Research

dpconline.org




sara day thomson

tw: @sdaythomson

# Digital Preservation Coalition

- 80+ Members
- Supporter Programme
- Capacity Building
- Good Practice and Standards
- Workforce Development
- Advocacy
- Community Building



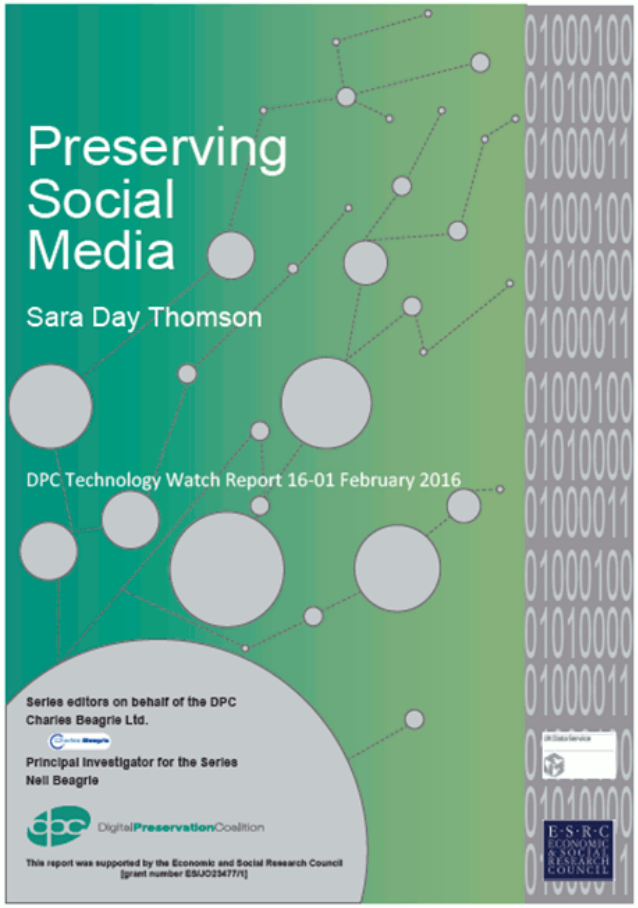






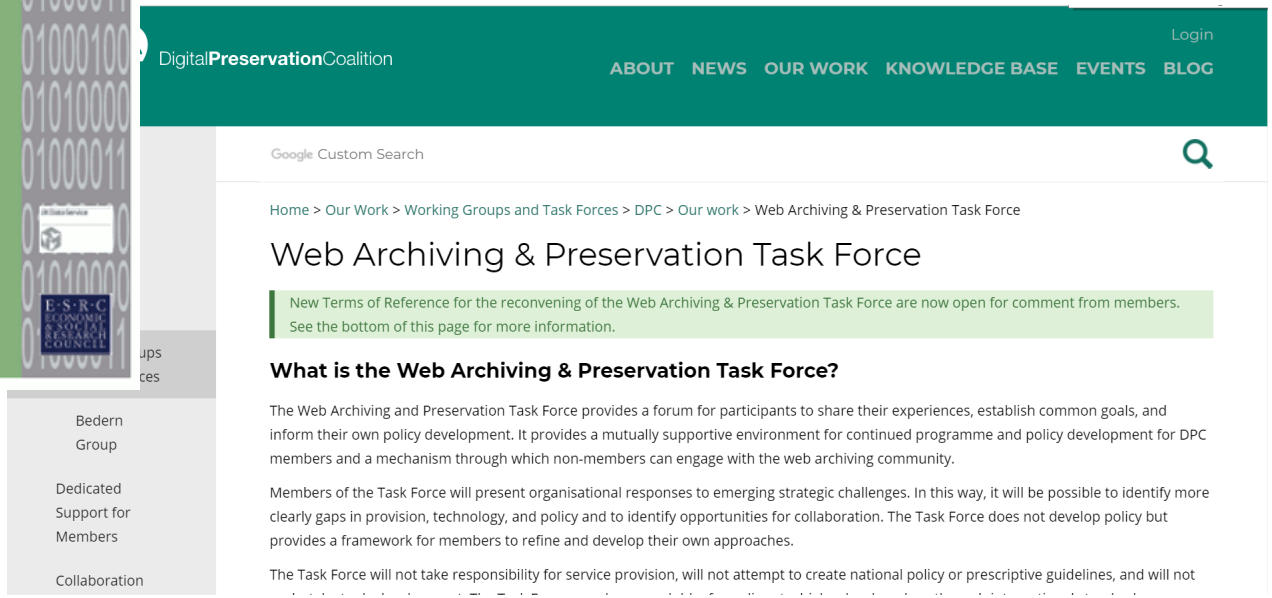


# Technology Watch Report




http://dx.doi.org/10.7207/twr16-01

## Web Archiving & Preservation Task Force (for DPC Members)



# Legally and ethically using, re-using, and preserving social media research data harvested from an API

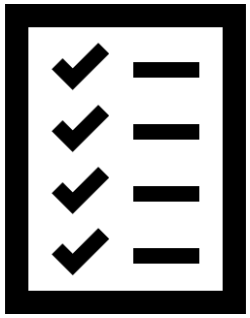
# Spectrum of Approaches

It doesn't have to be all or nothing…






DigitalPreservationCoalition



- What data is used and how?
- What are the legal & regulatory restrictions?
- What are the platform policy restrictions?
- What are the ethical implications?
- How will this affect users?

# What data is used and how?

DigitalPreservationCoalition

| Platform | Content & Metadata | Type of Analysis |
| --- | --- | --- |
| • Twitter<br>• Facebook<br>• LinkedIn<br>• FourSquare<br>• Flickr<br>• Etc | • Text<br>• Embedded URLs<br>• Images / Video<br>• Geolocation<br>• Number of followers<br>• Relationship to other users<br>• Etc | • Qualitative (e.g. language and linguistics, mental health)<br>• Quantitative (e.g. marketing analysis, transport data)<br>• Combination (e.g. sentiment analysis) |

# Legal & Regulatory Review

DigitalPreservationCoalition

## Privacy & Data Protection

- User generated content
- Collecting & storing data
- Sharing or exposing data
- Interacting with subjects

GDPR

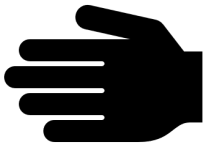
## Intellectual Property

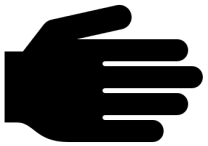
- Original content
- Embedded content, e.g. images or video

CDPA

## Open Data Policies

- Increase re-use
- Increase impact
- Funder mandates

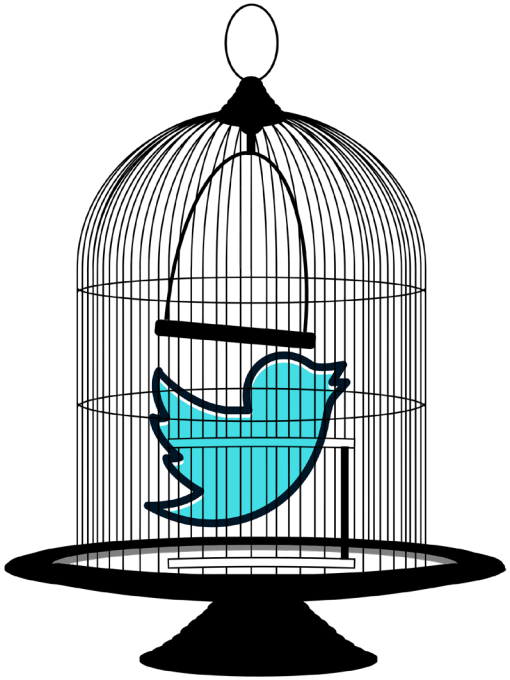




# Platform Restrictions

# Common Policies

- Terms of Service
- User Agreement
- Privacy Policy
- Terms of Use
- Developer Agreement or Developer Policy





Reproducibility



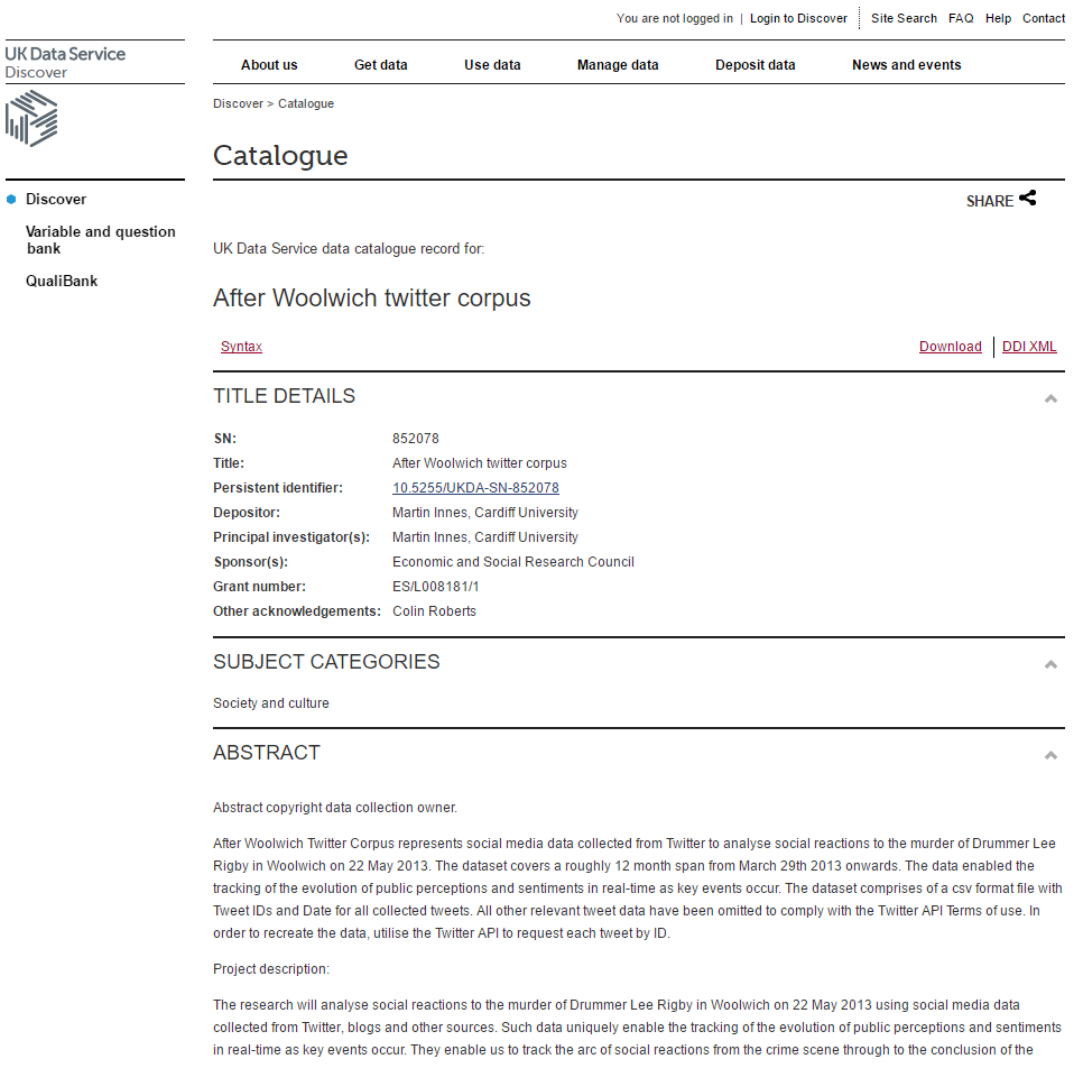
Validation

# Developer Agreement or Policy:

- Controls use of API
- Forbids sharing of data
- Forbids storing deleted data



# Example: Twitter Datasets in a Data Repository



You are not logged in | Login to Discover | Site Search FAQ Help Contact

UK Data Service
Discover

About us | Get data | Use data | Manage data | Deposit data | News and events

Discover > Catalogue

- Discover
- Variable and question bank
- QualiBank

## Catalogue

SHARE

UK Data Service data catalogue record for:

### After Woolwich twitter corpus

Syntax | Download | DDI XML

#### TITLE DETAILS

| | |
|---|---|
| **SN:** | 852078 |
| **Title:** | After Woolwich twitter corpus |
| **Persistent identifier:** | 10.5255/UKDA-SN-852078 |
| **Depositor:** | Martin Innes, Cardiff University |
| **Principal investigator(s):** | Martin Innes, Cardiff University |
| **Sponsor(s):** | Economic and Social Research Council |
| **Grant number:** | ES/L008181/1 |
| **Other acknowledgements:** | Colin Roberts |

#### SUBJECT CATEGORIES

Society and culture

#### ABSTRACT

Abstract copyright data collection owner.

After Woolwich Twitter Corpus represents social media data collected from Twitter to analyse social reactions to the murder of Drummer Lee Rigby in Woolwich on 22 May 2013. The dataset covers a roughly 12 month span from March 29th 2013 onwards. The data enabled the tracking of the evolution of public perceptions and sentiments in real-time as key events occur. The dataset comprises of a csv format file with Tweet IDs and Date for all collected tweets. All other relevant tweet data have been omitted to comply with the Twitter API Terms of use. In order to recreate the data, utilise the Twitter API to request each tweet by ID.

Project description:

The research will analyse social reactions to the murder of Drummer Lee Rigby in Woolwich on 22 May 2013 using social media data collected from Twitter, blogs and other sources. Such data uniquely enable the tracking of the evolution of public perceptions and sentiments in real-time as key events occur. They enable us to track the arc of social reactions from the crime scene through to the conclusion of the

# Sharing Tweet IDs





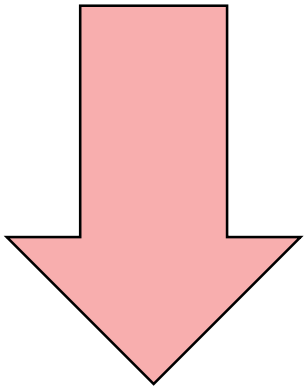
```
{
    "created_at": "Thu Apr 30 21:53:11 +0000 2015",
    "id": 593895901623496700,
    "id_str": "593895901623496704",
    "text": "This is a #test tweet @LoveforTestingT with an image. http://t.co/ZvgHovKZq4",
    "source": "<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client</a>",
    "truncated": false,
    "in_reply_to_status_id": null,
    "in_reply_to_status_id_str": null,
    "in_reply_to_user_id": null,
    "in_reply_to_user_id_str": null,
    "in_reply_to_screen_name": null,
    "user": {
        "id": 2993982541,
        "id_str": "2993982541",
        "name": "Test Demo",
        "screen_name": "jondee_test",
        "location": "Denver, CO",
        "url": null,
        "description": "this is a test account.",
        "protected": false,
        "verified": false,
        "followers_count": 2,
        "friends_count": 43,
        "listed_count": 0,
        "favourites_count": 0,
```



```
{
    "created_at": "Thu Apr 30 21:53:11 +0000 2015",
    "id": 593895901623496700,
    "id_str": "593895901623496704",
    "text": "This is a #test tweet @LoveforTestingT wit
```

# Policy & Ethical Review

# Ethical Requirements



Institutional Requirements e.g. Ethical Review Committee

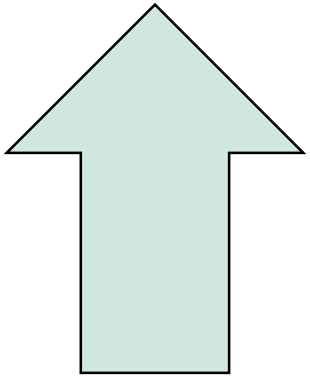
User Safety, Privacy & Awareness

Context & Relevance



Ethical Review of Data Collecting, Storage, and Analysis

Transparency of Process and Purpose



# User Views

NatCen Social Research
*Research Using Social Media: Users' Views* (2014)

Kelsey Beninger:

‘Research needs to be supported by an **inductive and flexible** approach to ethical thinking; what principles need to be considered in the **context of your study** and how can you think about these to ensure the **actions you decide to take** support an ethical study.’

'Social Media Users' Views on the Ethics of Social Media Research' in The SAGE Handbook of Social Media Research Methods, Eds Luke Sloan and Anabel Quan-Haase, London: 2017.

# 3 Suggestions…

1. Subject Recruitment:
   - state credentials
   - form a diverse and inclusive pool of users
   - clearly articulate purpose and aims

'Social Media Users' Views on the Ethics of Social Media Research' in The SAGE Handbook of Social Media Research Methods, Eds Luke Sloan and Anabel Quan-Haase, London: 2017.

# 3 Suggestions…

2. Protect participants:
   - ensure representativeness of findings
   - take into consideration the privacy risks created by the platform used

'Social Media Users' Views on the Ethics of Social Media Research' in The SAGE Handbook of Social Media Research Methods, Eds Luke Sloan and Anabel Quan-Haase, London: 2017.

# 3 Suggestions…

3. Protect identities & maintain trust:
   - set an example of transparency & honesty in reporting

'Social Media Users' Views on the Ethics of Social Media Research' in The SAGE Handbook of Social Media Research Methods, Eds Luke Sloan and Anabel Quan-Haase, London: 2017.

# Preserving Social Media



# social media is vulnerable to loss

*LinkedIn is not a storage service. You agree that we have no obligation to store, maintain or provide you a copy of any content or information that you or others provide.*

LinkedIn User Agreement

*The Twitter Entities shall not be liable for … any data loss.*

Twitter Terms of Service

# Bit Preservation

Even if immediate sharing is not possible….

…preserving the bits, metadata, and documentation is still critical to making social media content accessible in future.

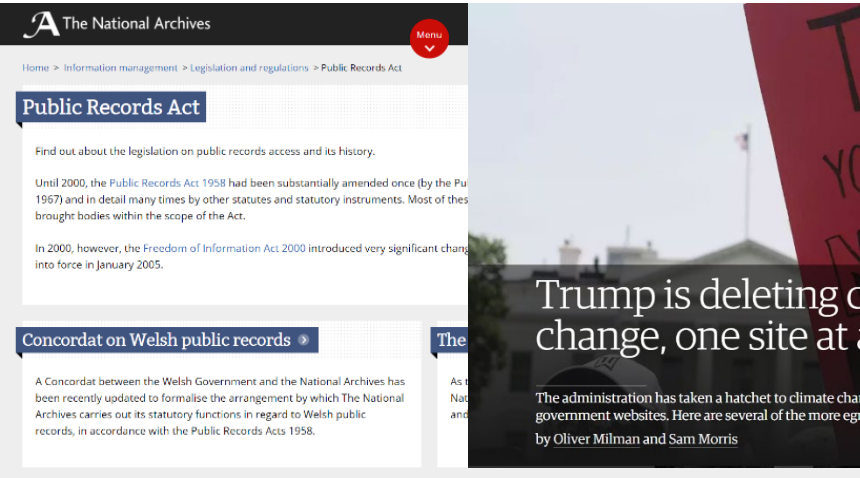


www.digitalbevaring.dk

# Advocacy!

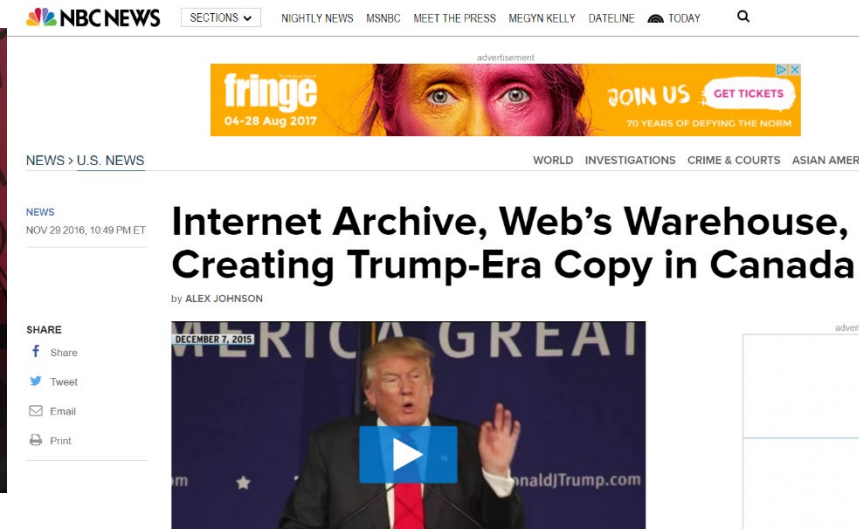


DigitalPreservationCoalition





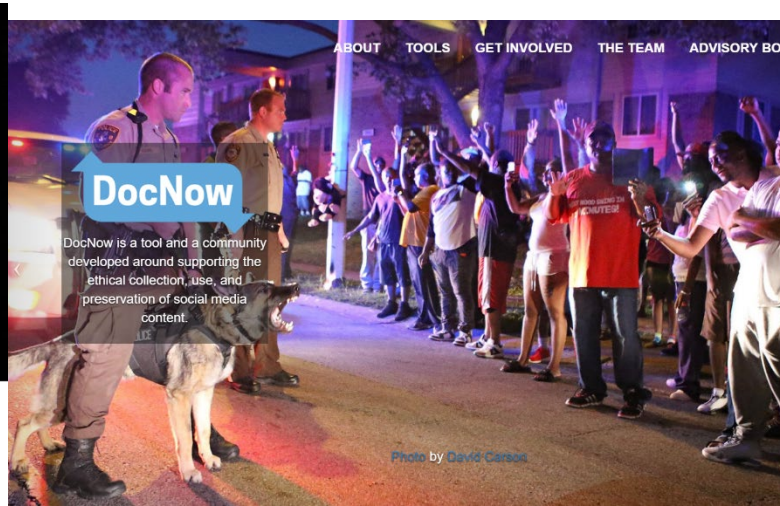










DigitalPreservationCoalition

# #thanks

