# Data Sharing

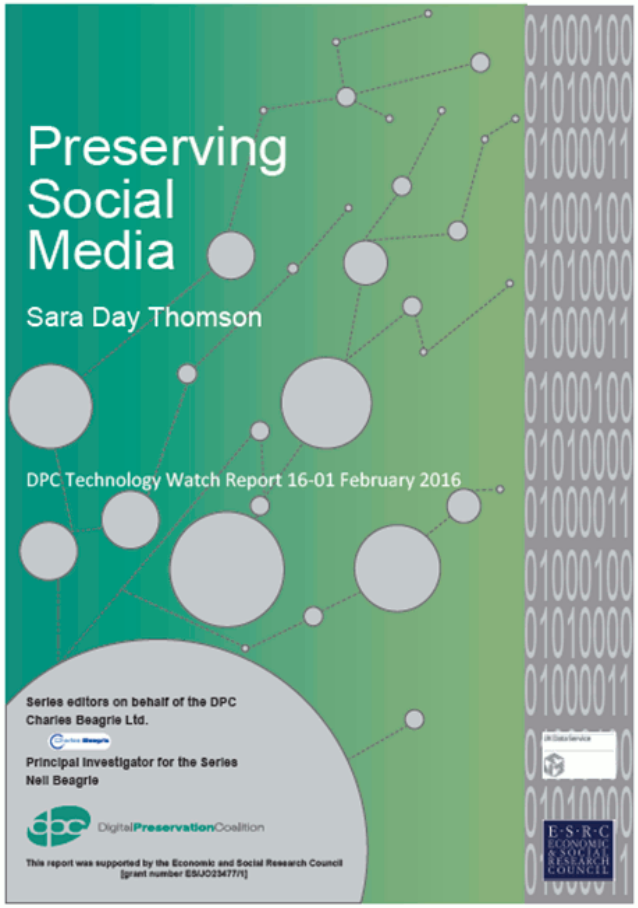
## digital preservation and data sharing

dpconline.org


sara day thomson

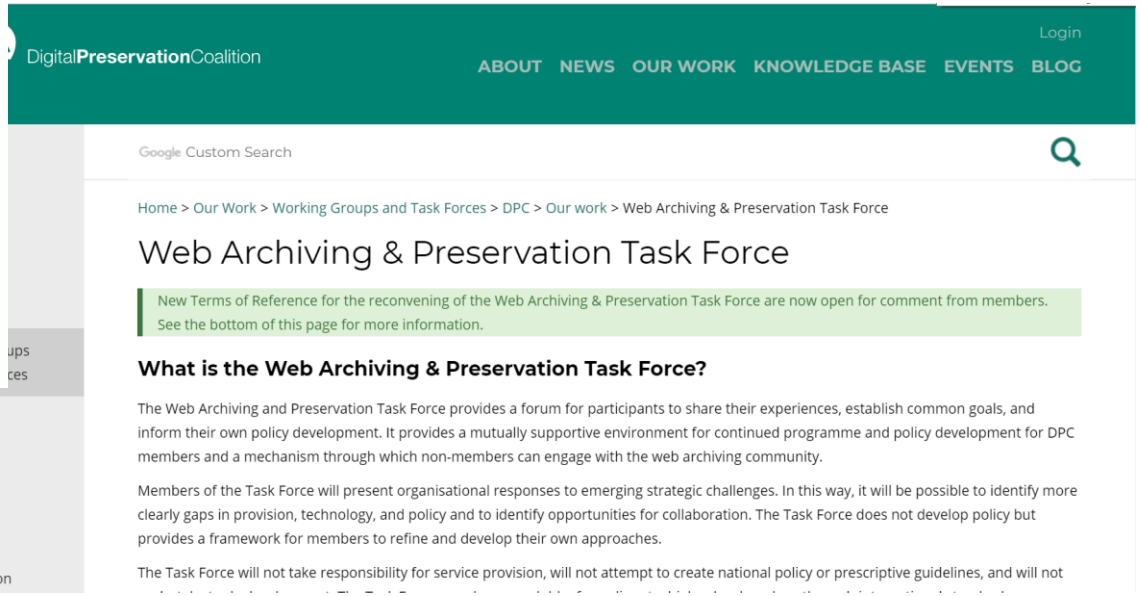
tw: @sdaythomson

# Technology Watch Report




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

## Web Archiving & Preservation Task Force



DigitalPreservationCoalition

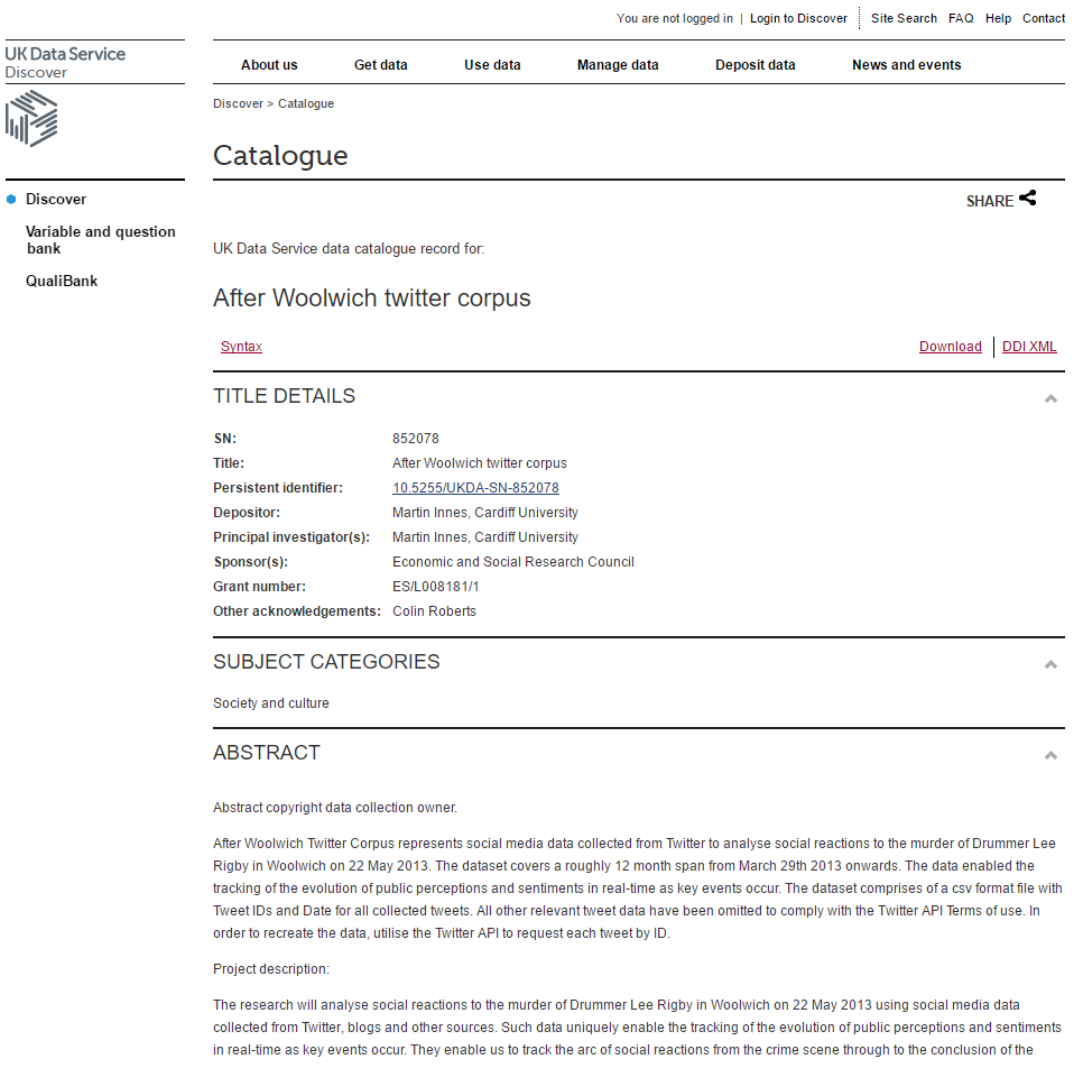
## Some Key Challenges

- sharing rich, useable data for reproducible research
- scale & volume
- ethics & user awareness
- lack of standards or best practice
- vulnerability of content

## Digital Preservation Approaches

- defining a Designated Community
- ‘re’-selection
- risk assessment & policy
- bit preservation & data security
- metadata & documentation
- advocacy

# Sharing 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

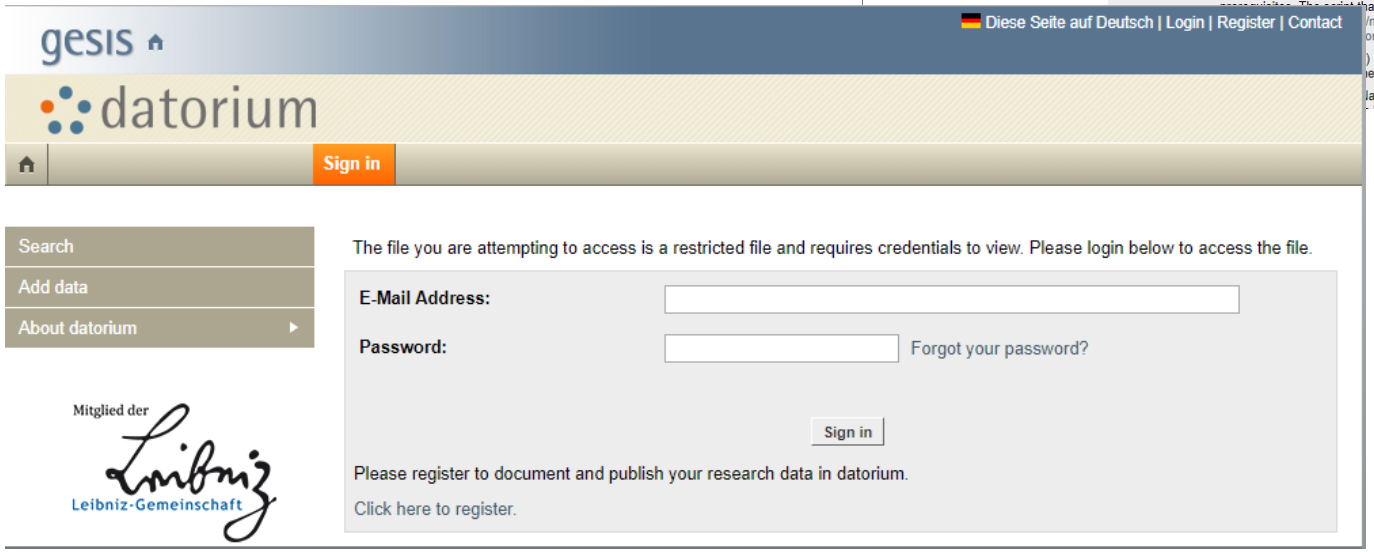
Abstract copyright data collection owner.

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.

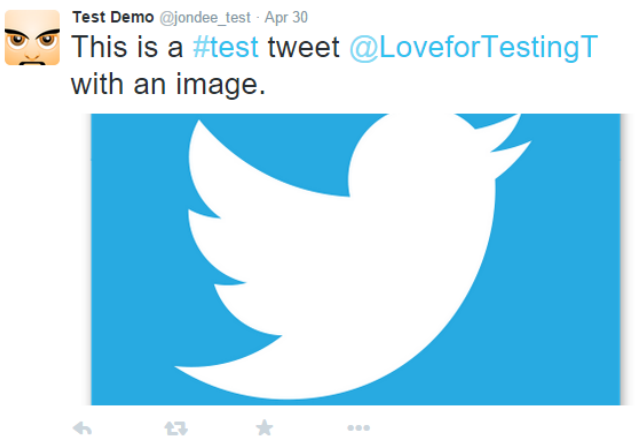
Project description:

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 Twitter Datasets in a Data Repository

# 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 with
```

# 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.

# Spectrum of Sharing Policies?

It doesn't have to be all or nothing…


# Risk Assessment & Policy Development

Develop sharing strategies based on th dataset or classificatio of datasets?

# Public Awareness

… about linked administrative data

‘At the beginning, **low awareness** of the uses of social research drove **scepticism** about its value.’

‘Later in the dialogues, when participants had **learned more** about the aims and methods of social research, they tended to be **more positive** about its value.’

*Dialogue on Data*
Ipsos MORI Social Research Institute

DigitalPreservationCoalition

# #thanks!