

preserving social media

applying principles of digital preservation
to social media archiving

dpconline.org



Digital **Preservation** Coalition

sara day thomson

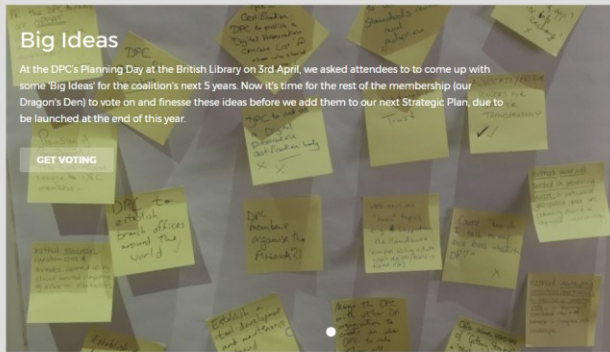
tw: @sdaythomson

Some Key Challenges

- capturing 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
- metadata & documentation
- advocacy



Big Ideas

At the DPC's Planning Day at the British Library on 3rd April, we asked attendees to come up with some 'Big Ideas' for the coalition's next 5 years. Now it's time for the rest of the membership (our Dragon's Den) to vote on and refine these ideas before we add them to our next Strategic Plan, due to be launched at the end of this year.

GET VOTING

We are building:

- A political and institutional climate responsive to the need for digital preservation
- Competent and responsive workforces ready to address the challenges of digital preservation
- Better tools, smarter processes and enhanced capacity in digital preservation
- Closer and more productive collaboration within and beyond the Coalition

JOIN US

Digital Preservation Handbook

The Handbook aims to identify good practice in creating, managing and preserving digital materials. By providing a strategic overview of the key issues, discussion and guidance on strategies and activities, and pointers to key projects and reports, the Handbook aims to provide guidance for institutions and individuals and a range of tools to help them identify and take appropriate actions.



Digital Preservation **Handbook**

Explore the Handbook

- Home
- Contents
- Introduction
- Digital preservation briefing
- Getting started
- Institutional strategies
- Organisational activities
- Technical solutions and tools
- Content-specific preservation
- Glossary

Digital Preservation Handbook



Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Welcome to the revised 2nd edition of the Digital Preservation Handbook. A key knowledge base for digital preservation, peer-reviewed and freely accessible to all.

"Overall the improvements to the Handbook make it, in my opinion, one of the more useful and flexible tools for identifying, understanding and getting to grips with practical approaches to the varying challenges of digital preservation. It uses approachable language, clear terminology and provides useful links to case studies and further reading which will be of benefit to students and practitioners alike." *Stefanie Davidson West Yorkshire Archive Service*

Digital information is increasingly important to our culture, knowledge base and economy. The Handbook, first compiled by Neil Beagrie and Maggie Jones in 2001, is maintained and updated by the DPC. This full revision (the 2nd Edition) has expanded and updated content to cover over 30 major sections (see [Contents](#)). The 2nd edition was compiled with input from 45 practitioners and experts in digital preservation under the direction of Neil Beagrie as managing editor and William Kilbride as chair of the Management and Advisory Boards. The Handbook provides an internationally authoritative and practical guide to the subject of managing digital resources over time and the issues in sustaining access to them. It will be of interest to all those involved in the creation and management of digital materials.



- New, updated version released summer 2016
- Written and reviewed by over 40 specialists in digital preservation
- Available on dpconline.org with additional embedded resources, such as short training videos



Who Is the Future?

Designated Community (OAIS Reference Model)

- define the main user group
- anticipate future user needs
- extract specialist knowledge to create metadata
- develop user requirements
- useful in lieu of established professional standards and best practice



UK Data Service

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

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



Digital Preservation Coalition

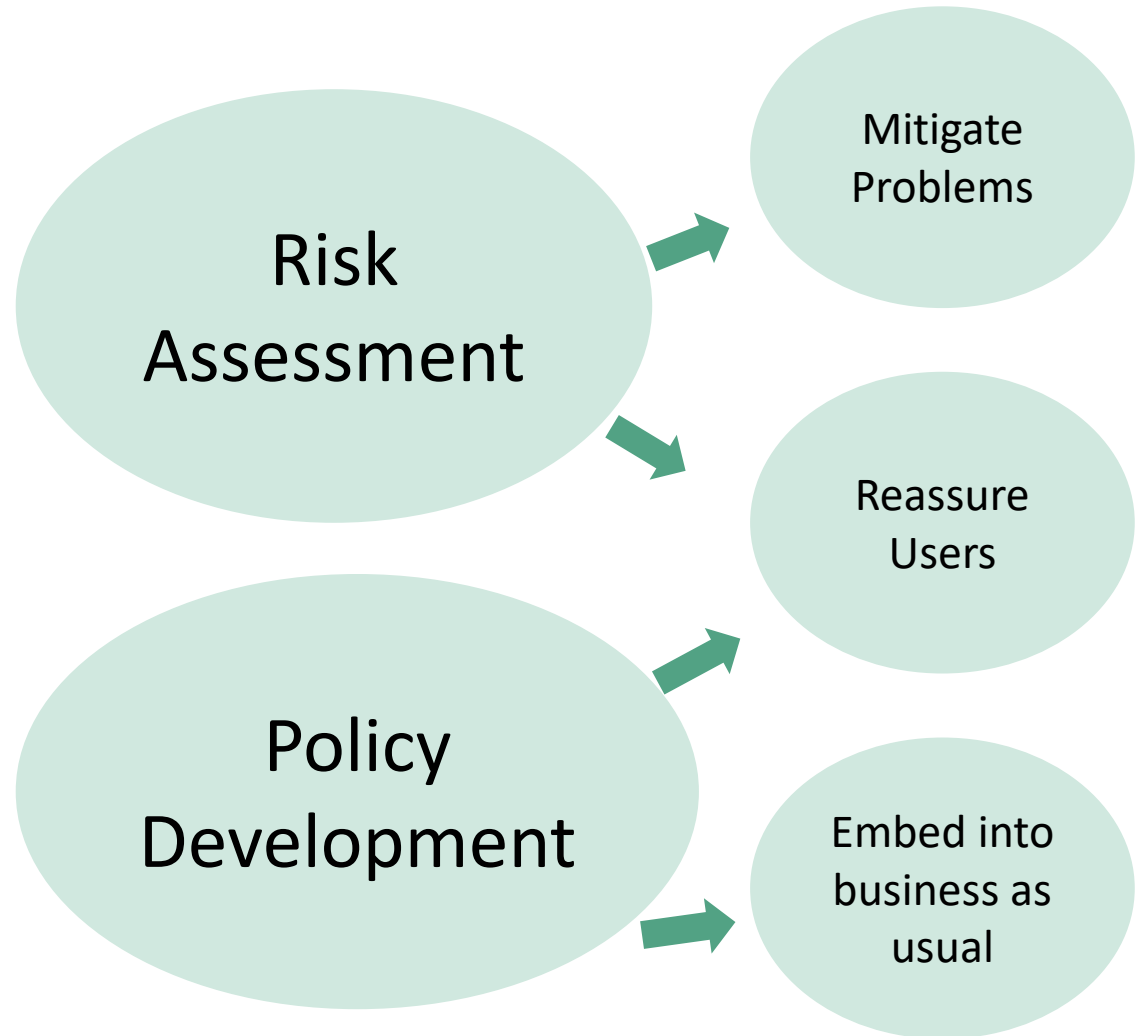
UK Web Archive

The screenshot shows a Twitter page for the account @libdem. At the top, there's a navigation bar with the Twitter logo and a sign-in button. Below that, a promotional banner for Liberal Democrats is visible. The main content area shows a tweet from @LibDemPress: "RT @paddyashdown to lead review of emergency aid http://j.mp/cAhRjF". The tweet has 23,243 retweets and 1,484 replies. Below the tweet, there are several replies, including one from @ThomasBrake and another from @Nick_Clegg. The right sidebar shows the profile information for @libdem, including their name, location (United Kingdom), and bio. It also lists several accounts they are following, such as @libdemppcs and @libdempeers.

Future-proofing

're'-selection

- limited data storage
- restrictions imposed by platform APIs
- personal and sensitive information
- implications of longer time scale: 10 years? 20? 50?





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

Future-telling: Bit Preservation

Even if immediate access is not possible....

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



www.digitalbevaring.dk

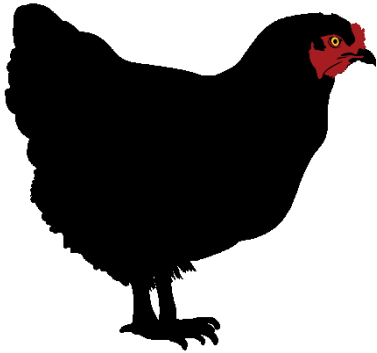
Public Awareness

... about linked administrative data

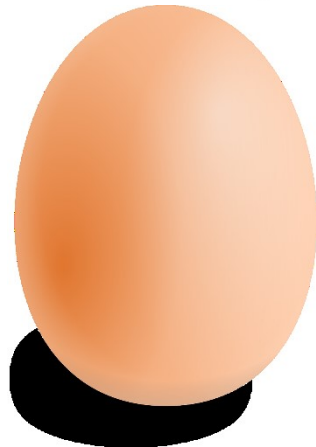
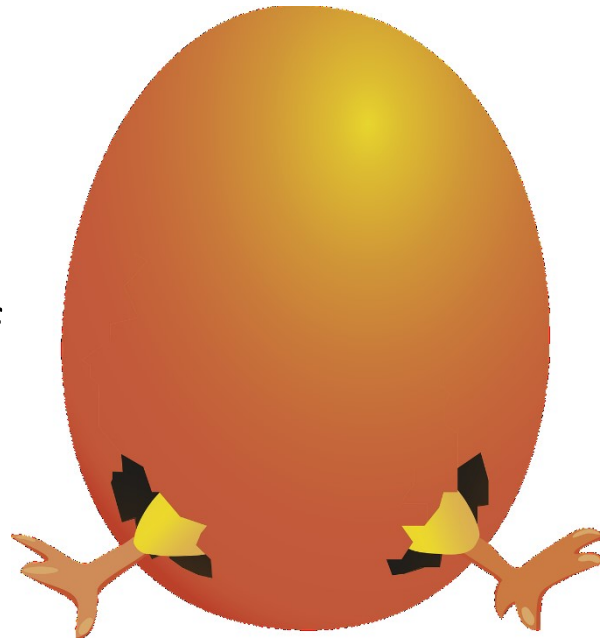
‘At the beginning, **low awareness** of the uses of social research drove **scepticism** about its value.’

‘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.’

Advocacy Conundrum



Robust, varied research
to demonstrate value of
social media as
research data



Better, more open
platform policies for
sharing social media
research data

Research has been possible *despite*
restrictive platform policies

Alignment with National Narratives

Transparent government & open data initiatives



Digital Preservation Coalition

NBC NEWS SECTIONS NIGHTLY NEWS MSNBC MEET THE PRESS MEGYN KELLY DATELINE TODAY



NEWS > U.S. NEWS

WORLD INVESTIGATIONS CRIME & COURTS ASIAN AMER

NEWS NOV 29 2016, 10:49 PM ET

Internet Archive, Web's Warehouse, Creating Trump-Era Copy in Canada

by ALEX JOHNSON



SHARE

- Share
- Tweet
- Email
- Print

The National Archives

Home > Information management > Legislation and regulations > Public Records Act

Public Records Act

Find out about the legislation on public records access and its history.

Until 2000, the Public Records Act 1958 had been substantially amended once (by the Public Records Act 1967) and in detail many times by other statutes and statutory instruments. Most of these amendments brought bodies within the scope of the Act.

In 2000, however, the Freedom of Information Act 2000 introduced very significant changes to the Act, which came into force in January 2005.

Concordat on Welsh public records

A Concordat between the Welsh Government and the National Archives has been recently updated to formalise the arrangement by which The National Archives carries out its statutory functions in regard to Welsh public records, in accordance with the Public Records Acts 1958.

Trump is deleting climate change, one site at a time

The administration has taken a hatchet to climate change language across government websites. Here are several of the more egregious examples by Oliver Milman and Sam Morris

Social movements reflected in social media

Muftah @MuftahOrg

Egyptian gov't can't diminish #Jan25 revolution: muftah.org/rememberingjan...
#انا_شارك_في_ثورة_يناير #I_participated_in_January_Revolution

12:10 PM - 20 Jan 2016

Remembering the Jan 25 Revolution

The Egyptian government is trying to dissuade protesters from gathering on January 25 to remember the revolution and call for muftah.org

Sky Views: Facebook's fake news threatens democracy

Some of the most viral news stories during the presidential campaign were entirely untrue.

Tom Cheshire, Technology Correspondent

Remember when Pope Francis endorsed Donald Trump for president?

ABOUT TOOLS GET INVOLVED THE TEAM ADVISORY BOARD

DocNow

DocNow is a tool and a community developed around supporting the ethical collection, use, and preservation of social media content.

Photo by David Carson

#thanks!

