

Table of Contents:

1. Introduction
2. Workshop agenda
3. Workshop activity summary
4. Appendix: Transcript of workshop post-it notes and sticker voting

Archiving Social Media Workshop Two

The workshop was organised by the British Library Web Archiving Team with support from Maja Maricevic, Head of Higher Education at the British Library on Monday 10th June 2019. The participants at the workshop were a mix of academic researchers, postgraduate research students and information professionals from the British Library, The National Archives, BBC and the Imperial War Museum.

The aim of this workshop was to review the requirements researchers have when working with social media data. The British Library are keen to get a closer understanding of the existing research projects where social media data was used or there was an intention to use it, but this was not possible. We are also interested to discuss re-use of data sets that could be created. This would help us towards having a more nuanced view of what would be useful for researchers working with social media data.

Details:

Date: 10/06/2019

Venue: Bronte Room, Knowledge Centre, British Library, London, NW1 2DB.

Time: 10:30 am registration/start at 11am to 4 pm

Agenda:

11:00 Introductions

11:10 **Tool demonstration:**

Collaborative Online Social Media Observatory ([COSMOS](#))

Mohamed Mostafa, [Cardiff University](#)

11:30 Q&A

11:40 **Academic research panel presentations:**

Chiara Bonacchi, [Stirling University](#)

Jane Winters, [School of Advanced Studies](#)

Bertram Vidgen, [The Alan Turing Institute](#)

12:20 Q&A

12:35 Lunch

13:20 **Using social media as a resource in PhD and Master Theses:**

Rhiannon Lewis, School of Advanced Studies

Fran Carpenedo, School of Advanced Studies

Trisha Harjani, formerly Warwick University (Pre-recorded Presentation)

14:10 Q&A

14:25 Comfort Break

14:35 Workshop Activity

15:45 Wrap Up

16:00 Close

Workshop Activity Summary

Each participant took a few minutes to write on a post-it note, one opportunity discussed during the talks that they could implement after this workshop and one challenge that could stop them from using archived social media in their research.

In their groups they discussed these positive and negative points about working with social media. They then placed these post-it notes in order of most achievable and most challenging. For some tables this was presented in a linear top to bottom row while other tables had many points they felt had equal weight and placed their post-it notes from left to right and other less important ones below this. When transcribing these notes they were put in order from left to right.

A spokesperson from each table summarised their discussions in two minutes.

In the final activity participants walked around the room with two different coloured stickers. They review what order other people have put on the tables and discuss their reasoning behind some decisions. They then place the gold star sticker on the post-it note that they thought was the most achievable thing they could take away from this workshop and the green dot sticker on the post-it note they felt had the most challenging aspect of using archived social media.

- The asterix symbol * represents the gold star sticker

- The black dot symbol ● represents the green dot sticker

A transcription of the post-it notes and how people individually voted is available in the appendix at the end of this document.

Analysis of the workshop data

There were a lot of common themes that came up in each group, these included reliability, ethics and challenges of capturing data.

One interesting point that can be seen by the way people voted is that something that one person put down as a positive point about working with social media one person put down as something that was challenging for them.

Also two people posted notes about the challenges that they face and as we can see in the table for Group A and Group C other participants saw these as achievable goals.

The top three points that participants felt were the most achievable for them were that ethics should be at the centre of all research projects, building a network with other researchers who are facing the same challenges and thirdly 'Lack of programming skills to analyse data' was voted the third most popular achievable most likely because there are more opportunities for researchers to develop these skills.

The top three points that participants felt were the most challenging for them were that there is a lack of diversity and fair representation within the social media data sets that are currently being collected while the second and third point raised was the limitation on what platforms data can be collected from.

| Most achievable | Voting Rank |
|--|----------------|
| Ethics should be at the centre of research and ethical guidelines/processes are (sort of) improving. | ***** (5) ● |
| Knowledge of a community of researchers with similar challenges | ***** (4) |
| Lack of programming skills to analyse data. | *** (3) |
| Researchers are interested in case reading of social content as well as 'big data' analysis. | * |
| Use of 'Hand Built' datasets. | * |
| Extracting sentiment | * |
| People's awareness that the social media data they work with has flaws. | * |
| Range of ways to approach analysis and gathering of Twitter data. | * |
| Network of people in social media research applied to humanities. | * |
| Sharing of tools, methods and knowledge by researchers (use of open source) | * |

| Most Challenging | Voting Rank |
|--|-----------------|
| Uneven demographic representation (social, economic, geographical, age). | ●●●●●●●●●● (10) |
| Not all social media can be collected/archived. | ●●● (3) |
| Unable to capture - Facebook and Instagram because of platform restrictions. | ●●● (3) |
| Lack of technical understanding and training. | ● |
| Terms of access for social media. | ● |
| Private ownership of data = Black Boxing and restricted access. | ● |
| Ethics should be at the centre of research and ethical guidelines/processes are (sort of) improving. | ***** (5) ● |

Appendix:

Transcription of the workshop post-it notes

Group A

| Achievable | Voting | Challenging | Voting |
|---|-----------|--|---------|
| Knowledge of a community of researchers with similar challenges | ***** (4) | Don't know what I can do with media data (eg. permissions, copyright). | |
| Interdisciplinary collaboration | | Not all social media can be collected/archived. | ●●●(3) |
| Useful data source: Doc Now Useful Tool: Hydrator | | Private ownership of data = Black Boxing and restricted access. | ● |
| | | Lack of programming skills to analyse data. | *** (3) |

Group B

| Achievable | Voting | Challenging | Voting |
|---|--------|--|--------|
| Range of ways to approach analysis and gathering of Twitter data. | * | No archived options or harvesting options. | |
| Network of people in social media research applied to humanities. | * | API restrictions | |
| Sharing of tools, methods and knowledge by researchers (use of open source) | * | Access to data mediated by service providers (mix of outputs, policies etc.). | |
| Topic modelling/Mallet | | Accessibility of Twitter perhaps leading to disproportionate amount of research produced on this platform. | |

Group C

| Achievable | Voting | Challenging | Voting |
|--|----------------|--|--------|
| Ethics should be at the centre of research and ethical guidelines/processes are (sort of) improving. | ***** (5) ● | Terms of access for social media. | ● |
| People's awareness that the social media data they work with has flaws. | * | Limit to access (1% of tweets). | |
| It is out there even if it is difficult to access and it gives us a perspective on society even if brief and incomplete. | | Lack of technical understanding and training. | ● |
| Some opportunities for self led development and learning. | | Social media tools are Black Boxes for social science researchers. | |
| Sentiment evaluator (especially with machine learning developments). | | That the data might be flawed. | |

Group D

| Achievable | Voting | Challenging | Voting |
|--|--------|---|-----------------|
| Lots of innovative and helpful research projects underway. | | Uneven demographic representation (social, economic, geographical, age). | ●●●●●●●●●● (10) |
| Availability of tools such as COSMOS and Vader. | | Facebook and Instagram not enabling scrapes/downloading. YouTube? - and comments. | |
| Variety - able to do | | The scale of collecting | |

| | | | |
|------------------------|---|--|--|
| research in new areas. | | of social media. It's huge - collecting takes months. | |
| Extracting sentiment | * | It feels like these projects are happening in isolation - need more input from libraries/archives. | |

Group E

| Achievable | Voting | Challenging | Voting |
|--|--------|---|--------|
| Availability of data (specifically in the case of Twitter). | | Understanding the sample of data available from social media platforms: How is the 1% selected? | |
| Potential archiving and reuse of product-generated datasets. | | Validity of data. | |
| Researchers are interested in case reading of social content as well as 'big data' analysis. | * | Reliability - Not a guaranteed representation of any grouping. | |
| Use of 'Hand Built' datasets. | * | Unable to capture - Facebook and Instagram because of platform restrictions. | 👍👍👍(3) |