

# Describing Research that uses Data from Social Media Platforms

Thank you for your interest in our survey; this survey forms part of the Provenance of Social Media Data project being conducted at the University of Aberdeen, UK. This project is funded by the Economic and Social Research Council, Social Media - Developing Understanding, Infrastructure & Engagement grant (ES/M001628/1).

The Provenance of Social Media Data project is exploring how understanding the provenance of data from social media platforms and its analysis in social science research could be a key role in increasing the transparency of such research. This, in turn, could support address growing concerns within the social sciences research community regarding the reliability, statistical validity, and generalisability of such research. Here, provenance is considered to be information about the entities (things), activities, and people involved in producing a piece of data or thing. For example, the provenance of a publication may include information about the entities included in the paper, such as graphs, tables, and text. Such information about, for example, a line graph may include details of the person that created it, when it was created, the tool used, and references to the data that is presented. Further information may also be provided about the data – for example, who collected it, when, where from, and how (including links survey forms that was used during data collection, if appropriate). The availability of such provenance information can increase the transparency of the research, which, in turn, could help others to understand the work that was undertaken, assess the analytical methods used, attempt to reproduce the work, and facilitate data reuse.

This survey aims to further our understanding of the current practices and attitudes towards describing the provenance of data collected from social media platforms and its analysis by researchers in the social sciences. This includes all forms of social media, such as Twitter, Facebook, Wikipedia, Quora, blogs, discussion forums, etc. Findings from this survey will influence the future work of this project around developing tools to support researchers increase the transport of research that uses social media data.

The survey is composed of three sections, and should take between 10-20 minutes to complete. If you have any queries about the survey or project, please contact Dr David Corsar (email [dcorsar@abdn.ac.uk](mailto:dcorsar@abdn.ac.uk), phone 01224 274562) or Dr Milan Markovic (email [milan.markovic@abdn.ac.uk](mailto:milan.markovic@abdn.ac.uk), phone 01224 274596).

\* Required

## Consent form for participation in the Provenance of Social Media Data project.

Please read the statements below and tick the box at the bottom to confirm that you have read and understood the statements, and upon doing so agree to participate in the project.



I confirm that the research project, Provenance of Social Media Data and this survey have been explained to me. I have had the opportunity to ask questions about the project and have had these answered satisfactorily.

I consent to the material I contribute being used to generate insights for the research project Provenance of Social Media Data.

I understand that my participation in this research is voluntary and that I may withdraw from the study at any time.

I consent to allow the fully anonymised data to be used for future publications and other scholarly means of disseminating the findings from the research project.

I understand that the information/data acquired will be securely stored by researchers, but that under the terms of the Economic & Social Research Council policy on research data, appropriately anonymised data may in future be made available to others.

**By ticking this box I confirm that I have read, understood, and agree to the above statements. \***

I confirm that I have read, understood, and agree to the above statements

**NEXT**

Never submit passwords through Google Forms.

---

This form was created inside of University of Aberdeen. [Report Abuse](#) - [Terms of Service](#) - [Additional Terms](#)

Google Forms



# Describing Research that uses Data from Social Media Platforms

\* Required

## Your Use of Social Media Data

What level of experience do you have using data from a social media platforms as part of your research? \*

- I have used/am currently using social media data as part of my research.
- I am aware of others using social media data as part of their research and may consider using it within mine.
- Neither of the above.

Did you attend either of the ESRC Social Media Training Workshops held at the University of Aberdeen? \*

- Yes, the Early Career Researcher workshop
- Yes, the PhD Student Research workshop
- No

[BACK](#)

[NEXT](#)

Never submit passwords through Google Forms.



# Describing Research that uses Data from Social Media Platforms

\* Required

## Use of Social Media Data

These questions aim to explore your existing/intended use of data from social media platforms within your research.

Briefly describe the research that you have conducted/are conducting using data from social media platforms. If you have used such data in more than one piece of research, please indicate this and base your responses to the questions in this section on the project that you are most familiar with. \*

Your answer

Which social media platforms have you obtained data from?  
Please check all that apply. \*

Twitter

Facebook



- LinkedIn
- Wikipedia
- Flickr
- Reddit
- Blogs
- Quora
- Forums
- Other:

How was data obtained from these social media platforms for your research? Please check all that apply. \*

- Using software tools developed by others (e.g. NodeXL, NCapture).
- Using software tools/scripts developed by myself.
- Manually, by copying from the webpage.
- By hiring external developers to obtain the data for me.
- Data was purchased from a private company (e.g. GNIP).
- Other:

How well do you feel you understand the process that was used to obtain data from social media platforms for your research? \*

- I have no understanding of the process.
- I understand parts of the process.



I understand the process completely.

Who was responsible for collecting data from social media platforms for your research? \*

Myself.

A colleague directly under my supervision (e.g. a PhD student, or Research Assistant).

Another colleague not directly under my supervision (e.g. a collaborator).

Other:

Who was responsible for preparing and analysing data from social media platforms during your research? \*

Myself.

A colleague directly under my supervision (e.g. a PhD student, or Research Assistant).

Another colleague not directly under my supervision (e.g. a collaborator).

A third party was contracted to prepare and analyse the data.

Other:

**BACK**

**NEXT**

Never submit passwords through Google Forms.

---

This form was created inside of University of Aberdeen. Report Abuse - Terms of Service - Additional Terms

Google Forms



# Describing Research that uses Data from Social Media Platforms

\* Required

## Describing the Collection and Use of Social Media Data

These questions aim to explore your attitude towards how the collection and use of data from social media platforms should be described in publications, data archives, etc.

Please indicate your level of agreement with the following statements: \*

Strongly disagree

Disagree

Neither agree or disagree

Agree

Strongly agree



Research using social media data

should be sufficiently described to allow audit of the collected data and analysis processes performed using it.

Describing the collection and use of social media data does not differ from describing the collection and use of other forms of data (e.g. from surveys).

I am aware of the information that should be provided to allow others to audit my use of social media data in research.

When describing my research, I ensure that sufficient information is always provided to allow others to audit my work..

When describing the collection of social media data, how important do you feel it is that the following information is provided? \*



Not important

Slightly important

Moderately Important

Important

Very important



Why the data was collected.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How the data was collected (e.g. listing any tools used).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When the data was collected.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which social media platforms data was collected from.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Who was responsible for collecting and managing the social media data.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The kind of data that was collected (e.g. Tweets, direct messages, images, videos, etc.).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Where the data was stored.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A partial description of the search/sampling criteria used.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A full description of the search/sampling criteria (including listing keywords, hashtags).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Details of any software used (including its name, version, and details of the operating system).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

When describing any data cleaning or preparation activities applied to social media data, how important do you feel it is that the following information is provided? \*



Not important      Slightly important      Moderately important      Important      Very important

Details of every activity that was performed.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The ordering of activities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Why each activity was performed.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Who performed each activity.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How each cleaning/preparation activity was done (for example, using a tool or manually).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

When describing analysis that used the social media data, how important do you feel it is that the following information is provided? \*

	Not important	Slightly important	Moderately important	Important	Very important
Details of every activity that was performed.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Why each activity was performed.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Who performed each activity.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How the analysis was conducted (for example, using a tool or manually).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please indicate the extent to which you agree with the statement: "A software tool that supports me to assess the extent to which descriptions of my research will allow others to fully understand and attempt to reproduce it, would be useful." \*

1                      2                      3                      4                      5



Strongly disagree      Strongly agree

Please indicate the extent to which you agree that you would use such a tool when preparing / submitting papers \*

1 2 3 4 5  
Strongly disagree      Strongly agree

**BACK**

**NEXT**

Never submit passwords through Google Forms.

---

This form was created inside of University of Aberdeen. [Report Abuse](#) - [Terms of Service](#) - [Additional Terms](#)

Google Forms



# Describing Research that uses Data from Social Media Platforms

\* Required

## Meta-Data

When depositing research outcomes, such as publications or datasets, in repositories, such as Research Fish or the UK Data Archive, you will typically be asked to provide meta-data about the outcome being deposited. Depending on the repository, some of this meta-data may be mandatory, for example providing a title, the name of the depositor, and the sponsor/funder; however, some meta-data is optional, such as subject categories, dates of data collection, or kind of data.

The following questions aim to explore your attitudes towards providing meta-data about research outcomes that are based on work that used data from social media platforms.

Have you deposited research data in the UK Data Archive? \*

- Yes
- No

Do/would you voluntarily provide additional meta-data about the research outcome beyond that which is mandatory? \*

- Yes
- No
- Sometimes

Which of the following reasons are influencing your ability to provide such additional meta-data? \*

- I would like to provide the additional meta-data but I do not have the time necessary to complete the form.



- I see no reason why I should provide the additional meta-data.
- There are too many obstacles in publishing such meta-data (e.g. querying databases to retrieve the necessary information, or ensuring the meta-data complies with licence restrictions associated with the data)
- The requested meta-data is not always relevant to the outcome(s) I deposit.
- The relevant information was not collected during the research to allow it to be provided as meta-data.
- I do not have the information needed to provide such meta-data
- None of the above
- Other:

Which of the following does/would motivate you to provide meta-data describing your research outcomes? \*

- Improving the overall accessibility and quality of research archives.
- Supporting research funders verify the proposed research outcomes



were delivered.

- Supporting other researchers find data to reuse during their work.
- Supporting the general public find research data to use for some purpose (e.g. developing new apps).
- Support other researchers understand my published research further by examining datasets and other outcomes associated with my research.
- None of the above.
- Other:

Would you be willing to provide the non-mandatory meta-data, if it could be generated semi-automatically by a software tool based on textual descriptions of your work? \*

- Yes
- No

**BACK**

**SUBMIT**

Never submit passwords through Google Forms.

---

This form was created inside of University of Aberdeen. [Report Abuse](#) - [Terms of Service](#) - [Additional Terms](#)

Google Forms

