

Codebook for ANES 2020 Social Media Study Restricted-Use Facebook Supplemental Data

Release: June 2, 2023

This codebook provides documentation for the restricted-use Facebook data component of the ANES 2020 Social Media Study survey dataset. The Facebook data are not a permanent archive; data are used by permission of Meta and are subject to removal upon expiration of a temporary research data agreement between Meta Platforms Inc. and Stanford University. The agreement is scheduled to expire no later than January 23, 2031.

The Facebook data have been prepared by Meta in consultation with ANES.

Funding acknowledgments: These materials are based on work supported by the National Science Foundation (www.nsf.gov) under grant number SES-1835022 to Stanford University (Shanto Iyengar, Principal Investigator) and SES-1835721 to the University of Michigan (Ted Brader, Principal Investigator). Any opinions, findings and conclusions, or recommendations expressed in these materials are those of the authors and do not necessarily reflect the views the National Science Foundation. These materials are also based on work supported by a gift from Facebook, Inc., to Stanford University.

These data may be cited as follows:

American National Election Studies. 2023. ANES 2020 Social Media Study Restricted-Use Facebook Supplemental Data [dataset and codebook]. University of Michigan and Stanford University. www.electionstudies.org

Table of Contents

- Table of Contents..... 2
- Consent, Privacy, Disclosures, and Disclaimers 3
- Merging the ANES 2020 Social Media Study with the Facebook Supplemental Data 5
- List of Facebook Supplemental Datasets 7
- Dataset: Facebook Daily Time Spent File: fct_time_spent.csv 10
- Dataset: URL Interactions File: fct_linkage_user_url_interactions.csv 11
- Dataset: URL Attributes (Interactions) File: dim_linkage_user_url_interactions.csv.csv 13
- Dataset: URL Shares File: fct_linkage_user_url_shares.csv 15
- Dataset: URL Attributes (Shares) File: dim_linkage_shared_url.csv 17
- Dataset: Political Ad Interactions File: fct_political_ads_exposure.csv 19
- Dataset: Political Ad Metadata File: dim_political_ad_metadata.csv 21
- Dataset: Posts with Candidate Mentions File: fct_candidate_post_mentioned.csv 23
- Glossary of Terms..... 25

Consent, Privacy, Disclosures, and Disclaimers

Consent

ANES obtained informed consent from survey participants to collect some of their Facebook information and share it with other researchers; see the ANES 2020 Social Media Study questionnaires at www.electionstudies.org. Out of 5897 survey respondents in the public use file, 4394 were self-reported Facebook users who were asked to consent to share Facebook information, and of these 4394, ANES obtained consent from 2739. Among these respondents, 1846 respondents were successfully linked to Facebook accounts that are included in the data release.

User Privacy

The research design for this dataset is built with consent and privacy in mind. For example, the design included obtaining explicit, informed consent from research participants for analyses of individual level data.

Privacy and security practices have also been implemented in the storage and processing of the data such as data access controls and isolated, privacy-protective storage of the data.

Protecting Participant Identity

Survey participants were assigned third-party IDs created by ANES to enable sharing of their Facebook activity without any direct identifiers. This process, which allowed ANES to securely link consenting study participants' survey responses to their Facebook data without personnel from Meta having access to participants' survey responses, consisted of the following:

1. ANES created a unique third-party ID for each participant.
2. ANES shared these third-party IDs with Meta via an encrypted inbound URL to Facebook for each participant who consented to share their Facebook activity.
3. Meta associated the IDs from ANES with an anonymized Facebook ID for each participant within Meta's infrastructure.
4. Meta stored the ANES third-party IDs on Meta's [Researcher Platform](#), hosted on AWS.
5. Data were deleted from the Researcher Platform once the data has been shared with ANES.

Disclosures

These metrics were created using internal Meta company data sources available in 2020. Data governance and sources may have changed since the study period and may not be representative of current operations at Meta. Due to issues, such as erroneously logging the data sources from which these metrics are created, these metrics may suffer from fluctuating data quality and incompleteness, which may lead to fluctuating accuracy. Within this codebook, we disclose known fluctuations where possible.

Researchers using this data are responsible for conducting standard and thorough data cleaning processes and are responsible for ensuring that their analyses are accurate. It is expected that researchers may find issues with the data when conducting their analysis. We encourage researchers to share any findings with us. This may include, but is not necessarily limited to, data quality, validity, or fidelity issues. Given the historical nature of the data and lapsed retention periods, we may not be able to fix the issue identified but, in such cases, will work to disclose them in an “Updates & Errata” report.

Disclaimers

Meta utilizes a variety of quality assurance measures to improve the accuracy, quality, and reliability of the data it shares for research purposes. However, given the volume of data released and the imperfection of any quality assurance process, inaccuracies persist. Meta makes no representation or warranty, express or implied, including without limitation, any warranties of fitness for a particular purpose or warranties as to the quality, accuracy or completeness of data or information. By accessing this data, researchers acknowledge that the data may contain some nonconformities, defects, or errors. Meta does not warrant that the data will meet the researcher's needs or expectations, that the use of the data will be uninterrupted, or that all nonconformities, defects, or errors can or will be corrected.

Merging the ANES 2020 Social Media Study with the Facebook Supplemental Data

The ANES 2020 Social Media Study (SMS) survey data and the Facebook supplemental (FS) log data can be merged at the respondent-level using the ANES unique case identifier variable “caseid”. Every respondent-level record in both the SMS data and the FS data is associated with a single “caseid” value. Some FS log data, describing interactions between respondents and specific content on Facebook (e.g., someone liking a URL on their feed), also include a unique identifier for Facebook content. The following synthetic examples illustrate how the SMS survey data and the FS data are conceptually linked. (Note: all identifiers in this example were generated for illustrative purposes and do not constitute actual data.) “FS log data” refer to records associated with specific respondents and “FS metadata” refer to data about elements of the log data, as illustrated below.

Example SMS survey data:

caseid	party_identification	age	...
10000	democrat	38	...
10001	republican	47	...
...

Example FS log data:

caseid	url_id	action	date	...
10000	5o135m8lsi9yc6	like	[date]	...
10001	ilb585fqppn3uuf	wow	[date]	...
...

Example FS metadata:

url_id	title	full_domain	...
5o135m8lsi9yc6	“Biden up in recent polls”	www.example.com/page1	...
ilb585fqppn3uuf	“Trump benefits from fundraising surge”	www.example.com/page2	...
...

As a basic template for merging the FS data with the SMS survey data, we suggest researchers start the following steps.

1. Determine which FS log data and FS metadata are relevant to the research question.
 - a. Note that some log data (e.g., time spent on Facebook) are not associated with any metadata.
2. Merge the relevant columns of the FS metadata with the relevant FS log data using metadata identifiers: “url_id” for URLs or “ad_library_id” for ads. Make sure to retain the “caseid” identifier as a column in the log data.
 - a. Ignore this step for log data not associated with any metadata.
3. Use the merged log data file, generate variables (columns) that can be aggregated up to the respondent level to answer the research question. For example, a variable that counts whether the log data represents exposure to a pro-Biden/pro-Trump political ad.
4. Aggregate the log data to the respondent-level for eventual merging with the SMS survey data.
5. Use the “caseid” identifier to merge (a subset of the) SMS survey data with the aggregated log data. The result is linked survey data and aggregated log data.
6. Analyze the final merged file to answer the research question.

Researchers should feel free to extend this basic template or develop their own data structures. For example, researchers might want to aggregate and then analyze data at the respondent-week level if they are interested in changes over time. However, we do not necessarily recommend analyzing data at the content level (e.g., URL-level or ad-level) since the FS data only captures a specific subset of Facebook content.

List of Facebook Supplemental Datasets

Dataset	File	Description
Daily Facebook Time Spent	fct_time_spent.csv	The amount of time each participant spent on Facebook in a given day.
URL Interactions	fct_linkage_user_url_interactions.csv	Data on participants' daily interactions with URLs from posts they saw or engaged with on Facebook during the study period. Interactions include content views and engagement such as comments, likes, reactions, link clicks, and shares.
URL Attributes (Interactions)	dim_linkage_user_url_interactions.csv	Characteristics of URLs from posts the participants saw or engaged with during the study period, as well as independent fact-checker ¹ ratings and details on the destination web page.
URL Shares	fct_linkage_user_url_shares.csv	The list of URLs that the participants posted during the study period, whether it was a reshare, and the date and time the posting happened.
URL Attributes (Shares)	dim_linkage_shared_urls.csv	Characteristics of URLs that the participants posted during the study period.
Political Ad Interactions	fct_political_ads_exposure.csv	The list of political ads participants saw or engaged with on Facebook during the study period. Includes impressions, reactions, comments, and shares, as well as the date the interaction occurred.
Political Ad	dim_political_ad_metadata	Select Ad Library metadata for political ads participants saw or engaged with on Facebook

¹ See <https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking>

Metadata	ata.csv	during the study period. Includes the content and source of each relevant ad.
Posts with Candidate Mentions	fct_candidate_post_mentioned.csv	Tallies the number of posts the participant posted with the name of major US 2020 presidential elections candidate in the text during a limited study period.

Aggregation and Filtering

Study period (except for posts with candidate mentions):

- Start: January 1, 2020
- End: November 3, 2020

Study period for posts with candidate mentions:

- Start: August 3, 2020
- End: November 3, 2020

URLs

The URLs included in this dataset are sourced from Meta’s internal datasets that have filtering in place to ensure that these have been publicly shared more than 100 times.

Data Types

The following types of platform activity data are available in this study dataset:

- **Daily time spent:** the amount of time, in minutes, a participant was actively engaging with Facebook during a specific time period or day.
- **Interactions:** Includes content views, engagement, and ads impressions.
 - **Content views:** the number of times content appeared on a participant’s screen on Facebook.
 - **Content engagement:** actions a participant initiated on organic content posted to Facebook. Includes likes, reactions, comments, shares, and link clicks.
 - **Comments:** the number of comments made on content posted to Facebook.
 - **Shares:** the number of times a piece of content was reshared.

- **Reactions:** the number of reactions on content. Includes Like, Love, Care, Haha, Wow, Sad, and Angry.
 - **Outbound link clicks:** The number of clicks or taps on links within posts that led to URL destinations off Facebook.
 - **Impressions:** the number of times an ad was on screen for the given participant, during a specific time period or day.
- **Posts with candidate mentions:** posts that include text with the names of US 2020 presidential candidates Biden, Trump, Pence and Harris.
- **Attributes:** data about the content participants posted, viewed or engaged with during the study period.
- **User political ad interactions:** participants interactions with political ads on Facebook during the study period. Interactions include impressions and engagement actions such as likes, reactions, comments and shares.

Dataset: Facebook Daily Time Spent

File: fct_time_spent.csv

This file contains data on the amount of time each participant spent on Facebook in a given day.

Data type

The amount of time, in minutes, a participant spent on Facebook while using either a web or app interface by date.

Data description

Each row corresponds to a participant, interface, and date combination.

Variable	Name	Description
Case Identifier	caseid	The unique respondent identifier.
Interface	interface	The primary type of interface the participant used to access Facebook: web or app. App includes native apps or mobile web versions of Facebook.
Daily time spent	time_spent	The number of minutes the participant actively spent on Facebook using the interface on the given day.
Date	ds	The date for which the participant's time spent is calculated. Reported as YYYY-MM-DD.

Dataset: URL Interactions

File: fct_linkage_user_url_interactions.csv

This file contains data on participants' daily interactions with URLs from posts they saw or engaged with on Facebook during the study period. Interactions include content views and engagement such as comments, likes, reactions, link clicks, and shares.

Data types

Interactions

- Content views: the number of times content appeared on a participant's screen on Facebook.
- Content engagement: types of actions a person initiated on non-sponsored content posted to Facebook. Include likes, reactions, comments, shares and link clicks.
 - Comments: comments made on content posted to Facebook.
 - Reactions: reactions made on content. Includes Like, Love, Care, Haha, Wow, Sad or Angry.
 - Outbound link clicks: clicks or taps on links within posts that led to URL destinations off Facebook.

Data description

Each row corresponds to a URL with the participant, type of interaction, and date-time the interaction happened. This dataset can be joined with the URL Attributes (Interactions) dataset using the "url_id" key.

Variable	Name	Description
Case Identifier	caseid	The unique respondent identifier.

URL ID	url_id	The unique alphanumeric string used to identify a URL the participant interacted with. A unique ID is generated for each URL string and internally randomized.
Event Time	event_time	The unix timestamp ² of when the interaction happened.
Interaction type	interaciton_type	Content views or engagement from the participant on posts with the URL. Includes views (“vpv”), likes, reactions, link clicks (“click”), and comments.
Date	ds	The date of when the interaction happened.

Disclosures

The Meta source table for this data has been deprecated which may lead to fluctuating accuracy.

This dataset only includes URLs posted to Facebook in a public post as a link attachment and posted or reshared more than 100 times on Facebook.

Timeline limitations

This dataset includes data from January 1, 2020 to November 3, 2020.

² Unix timestamps measure time in seconds elapsed since 00:00:00 UTC on January 1, 1970. Several open-source software packages can convert this into a familiar date or date-time variable.

Dataset: URL Attributes (Interactions)

File: dim_linkage_user_url_interactions.csv.csv

This file contains characteristics of URLs from posts the participants saw or engaged with during the study period, as well as independent fact-checker ratings (see <https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking>) and details on the destination web page.

Data type

Attributes: data on the content participants viewed or engaged with during the study period.

Data description

Each row corresponds to a URL that participants either viewed or engaged with. This dataset can be joined with the URL Interactions dataset using the “url_id” key. URLs have been processed by Meta to consolidate different web addresses that direct to the same destination and to remove potentially private and/or sensitive data.

Variable	Name	Description
URL ID	url_id	The unique alphanumeric string used to identify the URL.
URL	url_string	The complete web address.
Domain	full_domain	The name of the domain from the URL.
Title	share_title	The title of the web page the URL directs to, as defined by the author of the webpage.
Description	share_main_blurb	The description of the web page the URL directs to, as defined by the author of the webpage.
First date posted	first_post_time	The date and time the URL was first posted to Facebook by any account.

Fact-check rating	tpfc_rating	The rating independent fact-checkers assigned to the URL. Posts not sent to or rated by fact-checkers have a missing value. Ratings include true, false, prank generator, false headline or mixture, opinion, satire, not eligible, and not rated (i.e., missing).
Date of first fact-check	tpfc_first_fact_check	The date and time when the URL was first fact-checked, if at all. If the URL has not been fact checked, the value is missing.

Disclosures

The Meta source table for this data has been deprecated which may lead to fluctuating accuracy.

Only URLs posted to Facebook in a public post as a link attachment and shared more than 100 times on Facebook are included.

Dataset: URL Shares

File: fct_linkage_user_url_shares.csv

This file contains the list of URLs that the participants posted during the study period, whether each post was a reshare, and the date and time the posting happened.

Data type

Attributes: data on the content participants posted that contained a URL (i.e., shared) during the study period.

Data description

Each row corresponds to a URL with the participant, type of share, and date-time the share happened. This dataset can be joined with the URL Attributes (Shares) dataset using the "url_id" key.

Variable	Name	Description
Case Identifier	caseid	The unique respondent identifier.
URL ID	url_id	The unique alphanumeric string used to identify a URL the participant shared. A unique ID is generated for each URL string and internally randomized.
Event Time	event_time	The unix timestamp of when the share happened.
Share type	share_type	How the participant shared the URL to Facebook. Includes link or reshare.
Date	ds	The date when the share happened.

Disclosures

The Meta source table for this data has been deprecated which may lead to fluctuating accuracy.

The variable `share_type` is defined in the code as the content type. Content type is normally used to identify the format of individual pieces of content, such as posts. This does not affect the accuracy of the data for the purposes of this dataset.

Dataset: URL Attributes (Shares)

File: dim_linkage_shared_url.csv

This file contains characteristics of URLs that the participants posted during the study period.

Data type

Attributes: data on the content participants posted during the study period.

Data description

Each row corresponds to a URL that participants posted. This dataset can be joined with the URL Shares dataset using the “url_id” key. URLs have been processed to consolidate different web addresses that direct to the same destination and to remove potentially private and/or sensitive data.

Variable	Name	Description
URL ID	url_id	The unique alphanumeric string used to identify the URL.
URL	url_string	The complete web address.
Domain	full_domain	The name of the domain from the URL.
Title	share_title	The title of the web page the URL directs to, as defined by the author of the webpage.
Description	share_main_blurb	The description of the web page the URL directs to, as defined by the author of the webpage.
First date posted	first_post_time	The date and time the URL was first posted to Facebook by any account.

Fact-check rating	tpfc_rating	The rating independent fact-checkers assigned to the URL. Posts not sent to or rated by fact-checkers have a missing value. Ratings include true, false, prank generator, false headline or mixture, opinion, satire, not eligible, and not rated (i.e., missing).
Date of first fact-check	tpfc_first_fact_check	The date and time when the URL was first fact-checked, if at all. If the URL has not been fact checked, the value is missing.

Disclosures

The Meta source table for this data has been deprecated which may lead to fluctuating accuracy.

Only URLs posted to Facebook in a public post as a link attachment and shared more than 100 times on Facebook are included.

Dataset: Political Ad Interactions

File: fct_political_ads_exposure.csv

This file contains the list of political ads participants saw or engaged with on Facebook during the study period. It includes impressions, reactions, comments, and shares, as well as the date the interaction occurred.

For the purpose of this study, the following process was used to classify ads as political and reportable under the terms of respondent consent for revealing ads involving presidential candidates:

1. Meta identified Pages that ran “social issues, elections, or politics” (SIEP) ads from January 1, 2020, to November 3, 2020.
2. Meta shared those Pages with ANES.
3. ANES applied a keyword search for “Trump” and “Biden” to the Page names and categorized the resulting Pages as political in this context.
4. ANES shared the resulting list of Pages with Meta.
5. Meta applied the study’s inclusion criteria to include only ads run by those Pages, classified as political ads.

Data type

Impressions: the number of times a political ad was on screen for the given participant, during a specific time period or day.

Content Engagement: actions a person initiated on organic content posted to Facebook. Include likes, reactions, comments, shares and link clicks.

- **Comments:** the number of comments made on content posted to Facebook.
- **Shares:** the number of times a piece of content was reshared.
- **Reactions:** the number of reactions on content. Includes Like, Love, Care, Haha, Wow, Sad or Angry.
- **Link clicks:** The number of clicks or taps on links within posts that led to URL destinations off Facebook.

Data description

Each row corresponds to a participant engaging with or seeing a political ad on Facebook. Columns include the ad ID, the type of interaction, how many times it occurred, and the date.

Variable	Name	Description
Case Identifier	caseid	The unique respondent identifier.
Ad Library Identifier	ad_library_id	The unique string of numbers used to identify the ad the participant saw or engaged with. This ID is also used in the Ad Library Dataset.
Interaction Type	interaction_type	The type of interaction the participant had with the ad. Includes impressions or the engagement actions likes, reactions, link clicks, comments, and shares.
Interactions	interaction_value	The number of times the participant had that interaction with the ad on that day.
Date	ds	The day the data is reported for.

Disclosures

The Meta source table for this data has been deprecated which may lead to fluctuating accuracy.

Impressions

An impression is counted as the number of times an instance of a political ad is on screen for the first time. Example: If an ad is on screen and someone scrolls down, and then scrolls back up to the same ad, that counts as 1 impression. If an ad is on screen for someone 2 different times in a day, that counts as 2 impressions.

In a few cases when it can't be determined whether ads are on screen, such as on feature mobile phones, impressions are counted when ads are delivered to devices.

Missing values

Data for March 25, 2020 is not available for the political ads interactions dataset.

Dataset: Political Ad Metadata

File: dim_political_ad_metadata.csv

This file contains select Ad Library metadata for political ads participants saw or engaged with on Facebook during the study period. It includes the content and source of each relevant ad.

Data description

Each row corresponds to a political ad encountered by a participant. This dataset can be joined with the Political Ad Interactions dataset using the “ad_library_id” key. Refer to the ad library API at <https://www.facebook.com/ads/library/api/> for more details.

Variable	Name	Description
Ad Library Identifier	ad_library_id	The unique string of numbers used to identify the URL.
Creation Time	ad_creation_time	The unix timestamp of when someone created the ad. This is not the same time as when the ad ran.
Creative Body	ad_creative_body	A list of the text which displays in each unique ad card of the ad. Some ads run with multiple ad versions or carousel cards each with their own unique text.
Creative Link Caption	ad_creative_link_caption	A list of the captions which appear in the call to action section for each unique ad card of the ad. Some ads run with multiple ad versions or carousel cards each with their own unique text that appears in the link. (“Caption” refers to a web domain such as donalddjtrump.com or joe Biden.com.)

Creative Link Description	ad_creative_link_description	A list of text descriptions which appear in the call to action section for each unique ad card of the ad. Some ads run with multiple ad versions or carousel cards each with their own unique text describing the link.
Creative Link Title	ad_creative_link_title	A list of titles which appear in the call to action section for each unique ad card of the ad. Some ads run with multiple ad versions or carousel cards each with their own unique title text about the link.
Funding Entity	funding_entity	A string containing the name of the person, company, or entity that provided funding for the ad. Provided by the purchaser of the ad.

Dataset: Posts with Candidate Mentions

File: fct_candidate_post_mentioned.csv

This file contains tallies of the number of posts the participant posted with the name of major US 2020 presidential elections candidate in the text during a limited study period.

This dataset also contains data on the number of times US political candidates were mentioned in text participants posted to Facebook. Candidate names include Biden, Harris, Trump or Pence (see disclosures below).

Note that posts in this dataset were posted to Facebook between August 3rd, 2020 and November 3, 2020.

Data description

Each row corresponds to a participant who shared a post that mentioned a political candidate on a given date.

Variable	Name	Description
Case Identifier	caseid	The unique respondent identifier.

Posts	num_of_posts	The number of distinct posts mentioning candidates on the given day.
Candidate Mentions	likely_candidate_mentions	The names of the candidates mentioned in the posts and the number of times they were mentioned that day.
Date	ds	The date when the posts mentioning the candidate were posted to Facebook. Only content posted to Facebook between August 3, 2020 and November 3, 2020 is included.

Disclosures

Candidate names are identified by searching for keywords in the post text. We estimate 4% of posts in this dataset may have included text incorrectly attributed to candidates.

Keywords used include “biden”, “trump”, “kamala”, “pence”, “senator harris”, “sen. Harris”, and “kamala harris”. Only text matching those exact strings is attributed to a mention of a candidate name. Given the limitations of keyword search identification, candidate mentions may be missed for various reasons, including but not limited to the use of special characters or misspelling.

The keyword search was done on the text used when the post was first published to Facebook. Any edits made to the post after are not reflected in the data.

This dataset captures the data for a specific window of time. Posts are counted if posted to Facebook between August 3, 2020 and November 3, 2020 and available on Facebook when the data were captured in 2021.

Glossary of Terms

Term	Definition
Angry	A reaction represented by an angry emoji.
Candidate mentions	Posts that include text with the names of US political candidates Biden, Trump, Pence and Harris.
Comments	The number of comments on content.
Content	Types of published media that someone may share with their audience. This includes stories, posts, events, listings and more.
Engagement	Reactions, comments, shares and outbound link clicks on content.
Haha	A reaction represented by a laughing emoji.
Impressions	The number of times a political ad was on screen. This metric measures how often a political ads was on screen for a given audience.
Like	A reaction represented by a thumbs-up emoji.
Likes	The number of likes on content.
Outbound link clicks	The number of clicks or taps on links within posts that led to destinations or experiences off Facebook.
Love	A reaction represented by a heart emoji.

Term	Definition
Measurement period	<p>The date or date range the data is computing for.</p> <p>A majority of variables used in the study datasets aggregate over the date of January 1, 2020 to November 3, 2020. The measurement period for each variable is listed in the data dictionary.</p>
Reactions	<p>Reactions on posts on Facebook. The reactions button on a post allows people to share different reactions to its content: Like, Love, Haha, Wow, Sad or Angry</p>
Sad	<p>A reaction represented by a sad emoji.</p>
Shares	<p>The number of times a post was reshared.</p>
Political Ad	<p>An advertisement on Facebook classified as a political ad using keyword search of either “Biden” or “Trump.” This inclusion criteria was established by ANES and applied by Meta.</p>
URL	<p>The location of a webpage on the internet.</p>
Content views	<p>The number of times a piece of content appeared on a participant’s screen.</p>
Wow	<p>A reaction represented by a surprised emoji.</p>