User Guide for the Instagram Data Access Pilot for Well-being Research

May 2025
Version 2.0

Suggested Citation
Meta Platforms, Inc. (2025). *User Guide for the Instagram Data Access Pilot for Well-being Research*. (Version 2.0). Menlo Park, CA: Meta Platforms, Inc.

Table of contents

Executive summary

1. **Pilot program overview**
   - 1.1. Meta is partnering with the Center for Open Science (COS) on a pilot program to share certain Instagram data with a select group of academic researchers to study the social and emotional health of teens and young adults.
   - 1.2. This pilot program is based on a publishing model called "Registered Reports" where peer review is administered by an independent, academic Editorial Board and is conducted before data is collected and analyzed.
   - 1.3. This pilot program is designed to be independent. Researchers must design and recruit their own study participants using online or offline study designs, and neither the researchers nor study participants will receive financial payments or other compensation from Meta.
   - 1.4. The purpose of this User Guide is to detail what Instagram data can be made available as part of the pilot program, how to request additional data not listed in this User Guide, and how data will be shared by Meta so researchers with an approved Stage 1 Registered Report can merge the Instagram data with other study data they collect.
   - 1.5. See the "Pilot program overview" section for more details.

2. **Requesting Instagram data**
   - 2.1. As detailed in the Request for Proposals (RFP), Researchers will submit a Data Request Form after they submit their Pre-Proposal Form if they have been selected by COS and the Editorial Board to prepare a Stage 1 Registered Report. More details here.
   - 2.2. This User Guide provides a list of tables and variables that aims to be comprehensive for research on the topics listed in the RFP. Meta recognizes that there may be additional variables that researchers may wish to use for their Registered Report study. We invite researchers to request additional data on their Data Request Form, which Meta will review. More details here.
   - 2.3. Certain data types are not eligible for sharing as part of this pilot program, including but not limited to identifiers, such as profile or account names, demographic information, user generated content, ad content, or any inferred metrics. More details here.
   - 2.4. Meta will remain blind to the identities of researchers and the content of their study proposals when reviewing Data Request Forms. More details here.

3. **Collecting Instagram and other study data**

3.1. If researchers' Stage 1 Registered Report is accepted by COS and the Editorial Board, then the researchers and their institution will complete Meta's third-party onboarding requirements, sign Meta's Information Sharing Agreement, and coordinate with Meta for data collection and access. More details here.

3.2. Meta will provide approved researchers with a mechanism to link the data they collect with consented and authorized, user-level data for their study participants' Instagram account(s). This mechanism uses a URL that can be embedded in a data collection instrument to redirect study participants to Instagram where they can choose whether or not to authorize Meta to share data for their account(s). More details here.

4. **Accessing Instagram data**

4.1. Meta will share consented and authorized Instagram data with approved researchers as CSV files in a third-party cleanroom where the data can be joined with approved researchers' other study data and analyzed using standard statistical software (e.g. Excel, Stata, SPSS, R, Python). More details here.

Pilot program overview

Meta is partnering with the Center for Open Science (COS) on a pilot program to share certain Instagram data with a select group of academic researchers to study the social and emotional health of teens and young adults. Meta is excited to support global understandings of this important issue, and we are proud to continue to support public interest research.

This pilot program with COS is based on a publishing model popularized by the open science movement called "Registered Reports" where peer review is conducted before data collection and analysis. There will be four main steps in the submission and evaluation process for this pilot program: (1) Pre-Proposal Form; (2a) Stage 1 Registered Report and (2b) Data Request Form; (3) Data Collection; and (4) Stage 2 Registered Report. For greater detail on eligibility requirements and the submission and evaluation procedures at each step, researchers should consult the Request for Proposals (RFP) released by Meta and COS. The RFP also provides an anticipated timeline for the program, which researchers should consult when evaluating whether to submit a proposal.

This pilot program is designed to preserve the independence of scientific review and research processes. While Meta will evaluate and respond to researchers' data requests for their Registered Reports, Meta will not be involved in selecting which proposals move forward or which Registered Reports get approved for participating in the pilot. Instead, COS will evaluate researcher eligibility, and the Editorial Board they established will oversee the independent review and selection of Registered Report submissions. Meta will not be involved in study design or analysis. Neither the researchers nor study participants will receive financial payments or other compensation from Meta. Researchers must recruit their own study samples.

User Guide overview

This User Guide is intended for researchers who are interested in submitting a Registered Report as part of the pilot program. The purpose of this User Guide is to detail what Instagram data can be made available as part of the pilot program, and how that data will be shared by Meta so researchers with an approved Stage 1 Registered Report can merge the Instagram data with other study data they collect. This User Guide is accompanied by a Data Request Form which interested researchers can use to request the specific data they believe are needed to conduct their Registered Report study. This User Guide is complementary but separate from the codebook that will be provided for each approved

Registered Report study upon Meta's Instagram data delivery. Those codebooks will provide metadata for variables provided in the data delivery.

For the duration of the pilot, this User Guide might be updated with new or revised information, including information about the available data. This User Guide might become out of date after the pilot has concluded. Any changes to this User Guide will be described in the "Updates and errata" section and a new version will be released with a new version number. You can sign up for email alerts from COS here.

Requesting Instagram data from Meta

Submitting a Data Request Form

As detailed in the RFP, researchers will submit a Data Request Form if they have been selected by COS and the Editorial Board to prepare a Stage 1 Registered Report. The purpose of the Data Request Form is for researchers to request Instagram data for their Registered Report study.

The Data Request Form includes a list of tables and variables that have already been reviewed by Meta and can be made available as part of this pilot program, which this User Guide summarizes. This list of tables and variables aims to be comprehensive for research on the topics listed in the RFP. We recognize that there may be additional data that researchers require for their proposed studies that are not reflected in this list, or that Meta has not yet considered. Researchers can request additional data on their Data Request Form, which Meta will review.

Reviewing a Data Request Form

The Data Request Form will be reviewed by Meta in parallel with researchers' preparation and submission of their Stage 1 Registered Reports. Meta will be blind to both who submitted proposals and the content of their proposals when reviewing Data Request Forms. When completing the Data Request Form, researchers should avoid including information about their identities or hypotheses to ensure that Meta remains blind to the content of their proposals when evaluating data requests. COS will also thoroughly check and remove any remaining identifying information such as names or email addresses before sharing the Data Request Forms with Meta.

After completing its review of each Data Request Form, Meta will send a response to COS, who will then communicate the response to the Editorial Board and researcher. Meta's response will specify which of the requested data can be shared for their Registered Report study, and which of the requested data cannot be shared for their Registered Report study, if any. Meta may deny requests for additional data for reasons including but not limited to:

- if data are ineligible for this pilot as detailed in the "Data overview" section below;
- if data are unavailable;
- where data are restricted due to technical limitations;
- if giving access to the data could impact the operations of any of Meta's services, systems, or products, such as features Meta uses to promote safety, security, and integrity;
- for regulatory or legal reasons, such as conflicting with previous Meta commitments;
- if giving access to data could reveal confidential information such as trade secrets; or
- if giving access to data could compromise the legal rights, privacy, security, or integrity of other people or businesses.

Meta may suggest alternative data if the requested data is wholly unavailable or is unavailable in the format or granularity requested by the researcher.

Changing a Data Request Form

Researchers should carefully consider their requests as it will not be possible for researchers to change their requests or to submit a new request after submission of the Data Request Form prior to completion of Stage 1 Registered Report peer review.

Following completion of Stage 1 Registered Report peer review, if the Editorial Board, in agreement with the peer reviewers, recommend adding data from the list provided in this User Guide and/or removing previously requested data, then the Data Request Form can be updated in alignment with those recommendations. However, no additional data beyond the tables and variables listed in this User Guide can be requested at that time.

Collecting and verifying consent and authorization from study participants

Collecting participant consent

Researchers must recruit their own study participants and conduct their own study data collection or partner with an existing or forthcoming study such as a survey.[1] In addition to collecting account authorization using the method described in the section below, researchers will need to obtain relevant consent from study participants. When obtaining relevant consent for their study, each researcher's consent language for participants must adhere to the minimum transparency requirements set forth in Meta's Information Sharing Agreement. In particular, researchers must obtain:

1. **Study consent (or affirmative and explicit agreement) to participate in the study as a whole.** Additionally, for study participants under the age of 18 years old, researchers must first obtain consent (or agreement) from the participant's parent or legal guardian. Researchers are responsible for writing and collecting this consent (or agreement).

2. **Meta consent from study participants for Meta to share their Instagram data with the researcher(s)**. All participants for whom researchers want to link Instagram data must provide consent for Meta to share their Instagram data. For participants 13-17 years old, parental (or guardian) consent must first be obtained. For clarity, the collection of parental (or guardian) consent for participants 13-17 years old is in addition to the requirement to collect consent directly from such teens to share their Instagram data. This consent must be collected using the applicable consent(s) provided by Meta, including any Meta-translated versions for the languages used in a study. For reference, such language will be provided as part of Meta's Information Sharing Agreement to be executed by the researcher and Meta.[2]

[1] If partnering with an existing or forthcoming survey for the purposes of the pilot program, access to Instagram data must be stored in a separate file and access will continue to be restricted to the researcher and any other eligible researchers on their Registered Report study team for the purposes of conducting the research outlined in their approved Stage 1 Registered Report.

[2] If the researcher's study intends to collect Instagram data from participants 13-17 years old, in addition to Instagram data from their parents (or guardians) in order to merge such data together for analysis, Meta may require additional time to deliver the required Instagram data consent language.

In addition to obtaining relevant consents, Meta and each study must provide Frequently Asked Questions (FAQ) for study participants explaining how participants can request to withdraw from the study, who to contact with additional questions (including how to make data subject requests as required by applicable law), etc. The FAQ must be provided in a publicly accessible manner on the web in the languages used in the study. For more detailed information, see Meta's Information Sharing Agreement.

Collecting Instagram account authorization

In addition to study consent or agreement and Meta consent for data sharing, researchers need to collect authorization from participants for each Instagram account from which they choose to share their data. Meta will provide researchers with a mechanism to collect account authorization and to link the data they collect with consented and authorized, user-level Instagram data.[3]

If researchers' Stage 1 Registered Report is accepted by the Editorial Board, they will be directed to Meta to coordinate data collection and access. Before setting-up the Instagram account authorization though, researchers must provide evidence of approval from their Institutional Review Board (IRB) or equivalent, and researchers and their institution must complete Meta's third-party onboarding requirements and sign Meta's Information Sharing Agreement.

How Instagram account authorization works

To collect authorization, researchers will use URL(s) provided by Meta that can be included in their data collection instrument. These "inbound redirect URL(s)" will redirect participants to Instagram, where they can choose whether to authorize Meta to share the data for their Instagram account. After making their selection, participants will be redirected back to the data collection instrument via an "outbound redirect URL."

[3] If a researcher's study uses a mode other than an online survey for the majority of data collection, this portion of the study must still be conducted using a web-based questionnaire. For example, paper-and-pencil data collection might incorporate a QR code that links to a self-administered web questionnaire on a smartphone or tablet, and in-person data collection might incorporate use of a self-administered web questionnaire on a smartphone, tablet, or computer.

The technical infrastructure relies on these two URLs redirecting participants to and from the Instagram authorization flow:

1. The **inbound redirect URL** to Instagram authorization flow, which should be embedded in a button or link and will have a format similar to this: https://www.instagram.com/fbsurvey/confirm_user/?survey_fbid=XXX&id1=PLACEHOLDER&id2=PLACEHOLDER
2. The **outbound redirect URL** back to the study data collection instrument, which will have a format similar to this: https://demo.qualtrics.com/jfe/form/SV_XXXXXXXXXXX?id1=PLACEHOLDER&id2=PLACEHOLDER&status=completed[4]

The inbound redirect URL to the Instagram authorization flow will have up to four parameters that capture information about the researcher's data collection, which are detailed in Table 2 below.

Table 2. Inbound redirect URL parameters

Parameter	Description	Required or recommended	Static or dynamic	Provided by
survey_fbid	A unique number that identifies each inbound URL	Required	Static	Meta *Note: Shared with researchers via email*
id1	A unique, randomized identification number for participants in the researcher's study	Required	Dynamic	Researchers *Note: Passed to Meta via inbound redirect link; requires programming or manual input by researchers in the data collection instrument*
id2, id3	Page number(s), question identification number(s), or response identification number(s) used to specify a location in the researcher's data collection instrument that the researcher wants study participants to be directed to after they complete the authorization flow on Instagram	Recommended *Note: If the researcher does not use these identifier values, study participants will be redirected to the default page of the data collection instrument (e.g. the study consent)*	Dynamic	Researchers *Note: Shared with Meta via inbound redirect link*

[4] An example link using Qualtrics is provided here, but researchers can use any web questionnaire.

Box 1. Inbound redirect URL example

https://www.instagram.com/fbsurvey/confirm_user/?survey_fbid=123&id2=12&id1=456

With the example URL above, after study participant “456” completes the authorization flow on Instagram for study “123,” the participant is redirected back to a specific location in the data collection instrument based on id2. Note that researchers will need to specifically configure their data collection instrument to read the id2 parameter from the URL and use this to determine what participants are shown on redirect. For example, id2=12 will not *automatically* redirect participants to page 12 or question 12, but rather, researchers can program their data collection instrument to associate the value of id2 with a specific location in their instrument.

In the outbound redirect URLs back to the study instrument, Meta will pass up to four parameters, which are detailed in Table 3 below.

Table 3. Outbound redirect URL parameters

Parameter	Description	Static or dynamic	Provided by
id1	A unique, randomized identification number for participants in the researcher’s study	Dynamic	Meta *Note: Passed back by Meta via outbound redirect link*
id2, id3	Page number(s), question identification number(s), or response identification number(s) used to specify a location in the researcher’s data collection instrument that the researcher wants study participants to be directed to after they complete the authorization flow on Instagram	Dynamic	Meta *Note: Passed back by Meta via outbound redirect link*
status	A parameter shown if the participant agreed to authorize an account and not shown if the participant did not agree to authorize an account	Static	Meta *Note: Passed back by Meta via outbound redirect link, but researchers will need to do some programming to log the status*

Instagram account authorization flow options

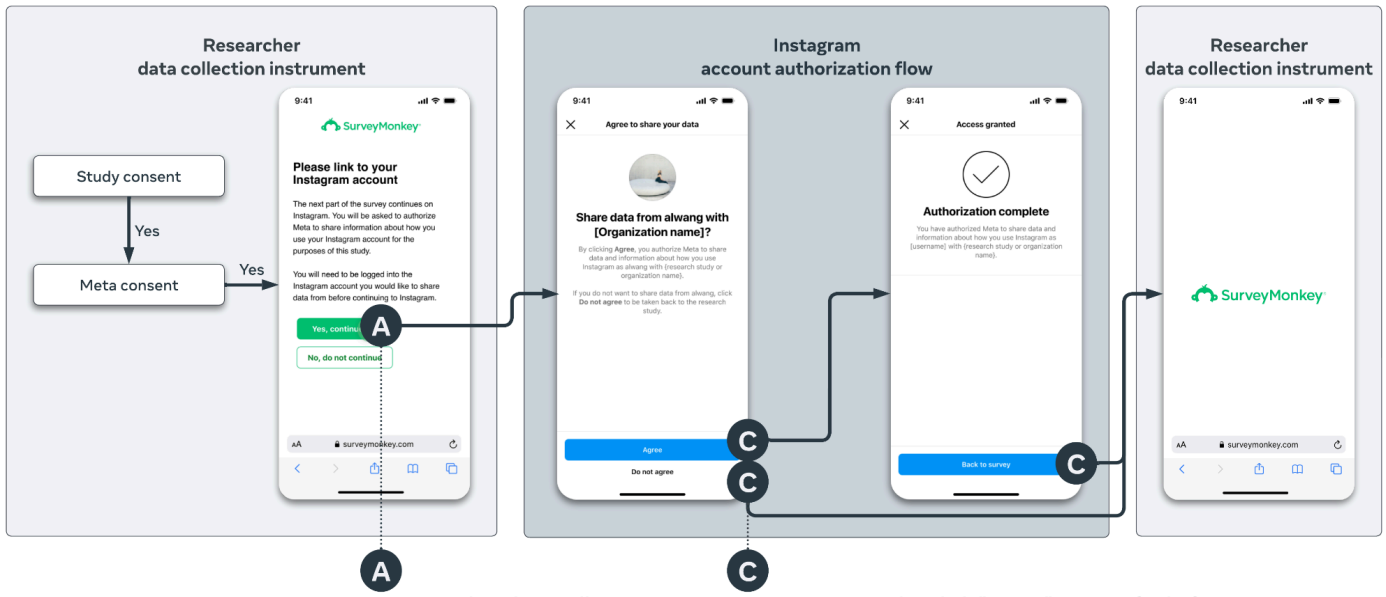
There are 3 Instagram account authorization flow options researchers can use in their studies: the auto login flow; the manual login flow; and the device choice flow. Each flow option has benefits and drawbacks that are unique to the considerations of each study, though the device choice or manual login flows will be most appropriate for most use cases. For example, if researchers are not sure whether participants will be on their own device, then the researchers should consider using the manual login or the device choice flow. However, the manual login flow may be preferable if researchers want participants

with multiple accounts to authorize a specific account rather than the account they happen to currently be logged into. These options are detailed in Table 1 and Figures 1 through 3 below.

Table 1. Instagram account authorization flow options

Option	Auto login	Manual login	Device choice
Description	All participants are redirected to the authorization screen for the account they are currently logged into. If not logged into an account, they will see the Instagram login screen instead and must enter their password. They must then go back to the researcher's study to get redirected again to the authorization screen.	All participants are redirected to the Instagram login screen and must enter their password.	Participants are redirected to a page where they can choose whether they are on their own device or someone else's. If on their own device, they will continue with the auto login flow. If on someone else's device, they will continue with manual login flow.
Consider using if...	• You're certain participants will be on their own device • You don't care which account participants authorize if they have more than one	• You aren't sure if participants will be on their own device • You're collecting authorization for multiple accounts from participants • You want participants to authorize a specific account if they have more than one • You're conducting a longitudinal study and want participants to authorize the same account as last time	• You aren't sure if participants will be on their own device • You don't care which account participants authorize if they have more than one
Key benefits	✓ Participants won't have to remember their password or choose an account to login with	✓ Helps prevent inaccurate authorizations for participants on someone else's device (e.g. teen's on their parent or guardian's device) ✓ More flexibility for researchers ✓ Easier for participants to choose which account to authorize	✓ Helps prevent inaccurate authorizations for participants on someone else's device (e.g. teen's on their parent or guardian's device) ✓ Participants won't have to remember their password or choose an account to login with if on their own device

Figure 1 . Auto login flow

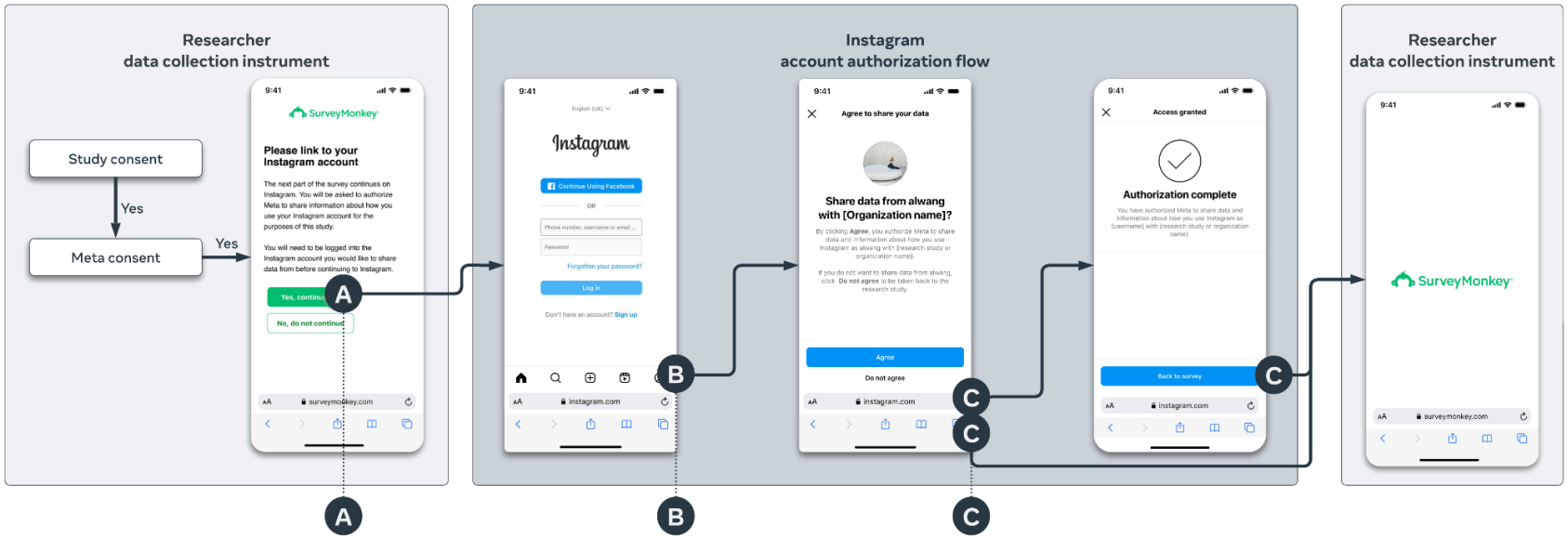


Researcher data collection instrument includes a Meta-provided URL, which redirects participants to Instagram for authorization of the account they are currently logged into

Participants who click "Agree" are notified of authorization, sent back to the researcher data collection instrument, and authorization is recorded by Meta

Participants who click "Do not agree" are sent back to the researcher data collection instrument and no authorization is recorded by Meta

Figure 2. Manual login flow



A

Researcher data collection instrument includes a Meta-provided URL, which redirects participants to Instagram for authorization

B

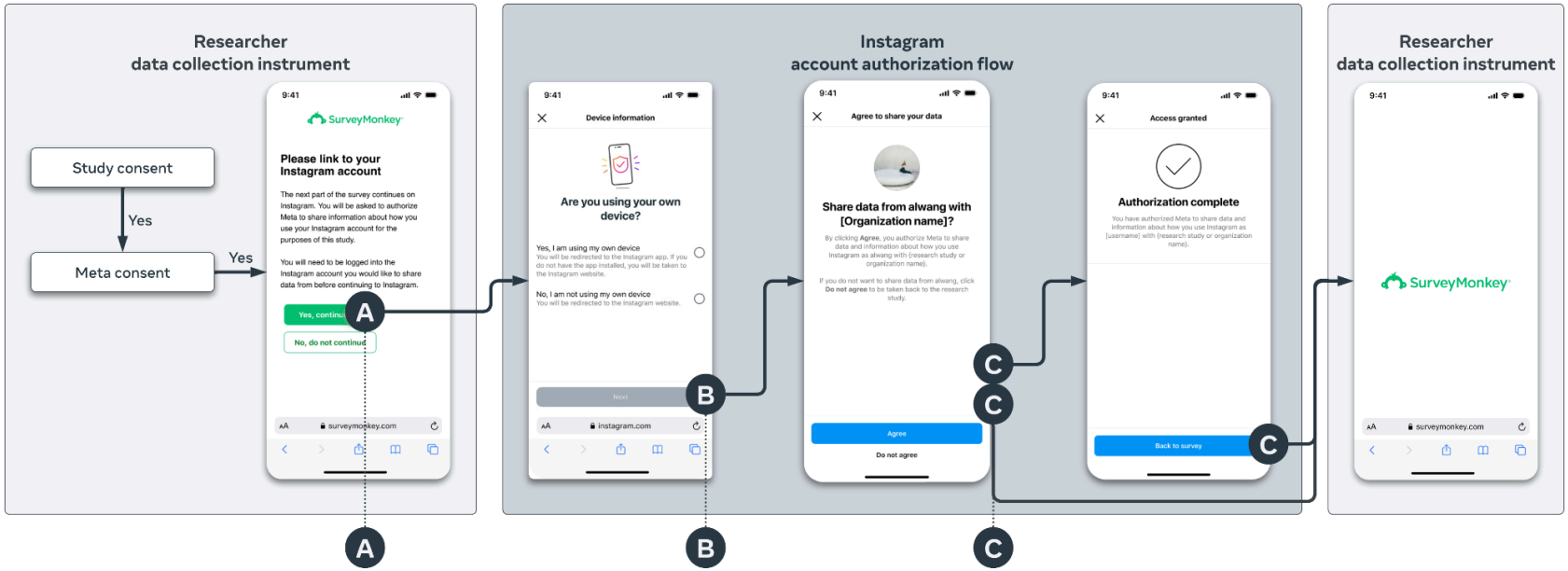
Participants who log in are directed to the appropriate authorization screen for that account

C

Participants who click "Agree" are notified of authorization, sent back to the researcher data collection instrument, and authorization is recorded by Meta

Participants who click "Do not agree" are sent back to the researcher data collection instrument and no authorization is recorded by Meta

Figure 3. Device choice flow for participant on their own device



A

Researcher data collection instrument includes a Meta-provided URL, which redirects participants to Instagram for authorization

B

Participants who select "Yes" and "Next" are directed to the authorization screen for the account they are currently logged into

C

Participants who click "Agree" are notified of authorization, sent back to the researcher data collection instrument, and authorization is recorded by Meta

Participants who click "Do not agree" are sent back to the researcher data collection instrument and no authorization is recorded by Meta

When designing their data collection instrument(s), researchers must incorporate Meta's provided consent addendum(s), and researchers must decide where and how in their survey to incorporate the inbound redirect URL(s) to the authorization flow on Instagram. At a minimum, this link to the authorization flow on Instagram must be included in a manner that will *ensure the link is only provided to participants eligible for Instagram data linking* (e.g. participants 13 years old or older and living in one of the 24 eligible countries).

Additionally, researchers need to decide how many inbound redirect URL(s) they will need. Each instance of consent in each data collection instrument and each wave of data collection need their own, unique inbound redirect URLs. And, if researchers are collecting authorization for more than one account per participant, researchers may also need more than one unique inbound redirect URL per data collection instrument and wave. This can be done in multiple ways, such as including a separate question for each additional account, or using a survey loop.

Researchers must tell Meta which flow option they want to use for each inbound redirect URL. Researchers may use a combination of flow options for their study. For example:

- A longitudinal study collecting authorization for one account from each participant may use the device choice flow in wave 1 and then the manual login flow for subsequent waves.
- A study collecting authorization for multiple accounts from each participant in any given wave might use the device choice flow or the auto login flow for the first account and then the manual login flow for any subsequent accounts authorized.

Setting-up Instagram account authorization

To set-up and use the authorization flow illustrated above, researchers must complete the following steps.

Step 1. Send study information to Meta

After designing their data collection instrument(s), researchers must provide Meta with the following information so that Meta can generate a unique URL that can be embedded within the researcher's data collection instrument:

- Name of their institution and/or study name that they would like referenced in the authorization flow on Instagram (see Figures 1 through 3 above)
- Approximate data collection period(s)
- Data collection instrument URL

- Data collection instrument label (e.g. ‘parent survey’ or ‘survey version 1’)
- Data collection wave (e.g. 1, 2, 3)
- How many inbound redirect URL(s) are needed for that wave, and the authorization flow type for each (i.e. auto login, manual login, or device choice)
- [Optional] Specifications for the id2 and id3 parameters defined in Tables 2 and 3

Step 2. Meta will provide the inbound redirect URL(s)

Meta will then provide the inbound redirect URL(s) with placeholder(s) for the id1 parameter, as well as the id2, and id3 parameters if researchers want them.

Step 3. Program the placeholder values and logging for the redirect URLs

After receiving the inbound redirect URL(s) with placeholder values from Meta, researchers will then need to program:

- id1 parameter in the inbound redirect URL
- [Optional] id2 and id3 parameters defined in Tables 2 and 3
- [Optional] Logging for the authorization status parameter passed via the data collection instrument redirect URL
- [Optional] Logic to allow authorization for multiple accounts

At a minimum, when directing participants from their data collection instrument(s) to Instagram, researchers must replace the “id1” placeholder in the inbound redirect URL with a unique identifier for each participant (i.e. their participant_id). If using id2 and id3 parameters, researchers must replace these placeholders with a unique identifier for the page number, question identification number, or other value being used to direct study participants to the correct location in their data collection instrument after they complete the authorization flow on Instagram. There are a number of ways to populate the id1 and id2 values, and the appropriate choice will depend on researchers’ data collection methods / instruments.

The following scenarios are examples of how programming the id1 values might work:

- **For programmatic data collection:** Write custom code or use survey programming to automatically insert the correct unique value for this parameter into the inbound redirect URL. For example, if using Qualtrics, researchers might use embedded data fields to program their data collect instruction to automatically populate the id1 and id2 parameters.
- **For manual data collection:** Generate a list of links populated with participant IDs and share the corresponding link with each participant directly (e.g. via email).

Testing the Instagram account authorization

After programming their data collection instrument and inbound and outbound redirect URL parameters, researchers can test their questionnaire by asking Meta to turn on the "test mode" for their study. When in test mode, researchers will see a message at the bottom of the screen: "You're in test mode." Researchers will not be able to provide authorization to Meta for data sharing while in test mode. To begin collecting authorization from study participants, researchers must request Meta to switch the link from test mode to "public."

Instagram account authorization best practices

1. **Account for drop-off in the sample size:** Participants may choose not to authorize an account, even if they previously consented to share their data. Additionally, participants may inadvertently drop out of the authorization and/or data collection flow. For example, if study participants do not click "Agree" or "Do not agree," then they might have trouble navigating back to the researcher's data collection instrument, causing their authorization to drop-off.
2. **Account for time limits:** Study participants may have set time limits for social media accounts or apps in their account settings and/or on their device(s) such as in their iOS or Android settings or through a third-party supervision app. If participants with a time limit exceed that limit before clicking on the redirect link to the authorization flow on Instagram, they will see their standard time limit alert, and they will need to either (a) turn off, extend, or override that time limit or (b) retry on a different day before the time limit is exceeded in order to access the authorization flow.
3. **Multiple accounts:** Carefully consider whether and how the study population uses multiple accounts when drafting the redirect page(s) and deciding how many redirect page(s) to include or which Instagram authorization flow(s) to use. Also note that for the purposes of this pilot, Meta will only share data for multiple accounts if they are linked to the first eligible account the study participant authorized in the Meta Accounts Center at the time the authorization was made. If the participant links another account to the Meta Accounts Center at a later time (e.g. in a later wave of a longitudinal study), it will not be provided in the data delivery file(s), even if authorized by the participant. This helps to ensure that the authorized accounts belong to the study participant and not someone else who has not consented to the study but who rather found the URL by other means (e.g. if a study participant retrieves and shares the URL). See a summary of this and potential other sources of data loss in the "Data merge and quality checks" section.

4. **Data collection mode:** Avoid recruitment or data collection methods that require participants to complete the authorization flow on a device other than their own. Otherwise, they may authorize an account that is not their own. For example, researchers may want to avoid "hot hand-off" recruitment where a parent or guardian consents to their teen participating in the study before then handing the device directly to that teen. If researchers cannot avoid these sorts of recruitment or data collection methods, then they may wish to consider using the device choice or manual login flows only and asking a question at the end of the survey about whose account they logged into and authorized.

Consent verification

At the conclusion of the researcher's study data collection, prior to Meta sharing Instagram data, researchers must certify to Meta the list of "participant IDs" that correspond to research participants who, according to their records, have valid, active agreement:

- **Valid agreement to participate:** Researchers confirm that participants (a) consented (or provided affirmative and explicit agreement) to participate in the study and (b) consented to Meta sharing their Instagram data with the researcher. For 13-17-year-old participants, a parent or guardian must have provided consent (or affirmative and explicit agreement) prior to the teen's recruitment and consent.
- **Active agreement to participate:** For participants 18 years or older, this means that they have not withdrawn from the study or requested that their study data be deleted. For 13-17-year-old participants, neither they nor their parents or guardians have withdrawn from the study or requested that their study data be deleted.

To certify this list, researchers must send Meta a CSV file with:

- The **participant ID** for all respondents who have valid, active agreement.
- The **timestamp** showing when each respondent provided affirmative consent to Meta's consent addendum. For 13-17-year-old participants, the CSV file must include the timestamp for their parent or guardian's affirmative consent, as well as the timestamp for their own affirmative consent, to the applicable Meta's consent addendum. This timestamp can be logged automatically in the case of a web or computer-assisted survey or logged manually at the time consent is recorded by

the researcher if using a mode of data collection that would not allow automatic logging.[5]

For studies collecting data from both 13-17-year-old participants and participants 18 years old or older, researchers must send separate CSVs, one for 13-17-year-old participants and one for participants 18 years or older:

- For the 13-17-year-old CSV, the file should be named "**13_17_consent.csv**" and should have only three columns: "**participant_id**"; (2) "**parent_consent**" timestamps for parents' (or guardians') consent, and (3) "**teen_consent**" timestamps for the teens' consent.
- For the 18 years or older CSV, the file should be named "**18_or_older_consent.csv**" and should have only two columns: (1) "**participant_id**"; (2) "**adult_consent**" timestamps for the adults' consent.

All timestamps should be in Unix time (i.e. the number of non-leap seconds that have elapsed since 00:00:00 UTC on January 1, 1970).

Instagram data delivery

Meta will merge the consent certification CSV(s) with data logged by the Instagram authorization flow, conduct final data processing and file quality checks (see the "Data merge and quality checks" section), and deliver the Instagram data to a third-party cleanroom (see the "Accessing Instagram data" section). The Instagram data delivered to the third-party cleanroom will be attached to the participant IDs provided by the researchers (a) in the authorization flow's inbound URL (see Box 1) and (b) in the consent certification CSV(s). This will allow researchers to merge their other study data with the Instagram data for the purpose of conducting the research outlined in their Stage 1 Registered Report. The data for each study will be accompanied by its own codebook with variable specifications and other information, such as an aggregate summary of the participant ID(s) and/or account(s) dropped.

Accessing Instagram data

Meta will provide access to the Instagram data for the pilot studies in a secure cleanroom environment hosted by Secure Data Access Centre (CASD). CASD employs a secure

[5] For example, the researcher might add a field for date and time on the physical consent form(s) if conducting a face-to-face interview or paper-and-pencil questionnaire.

equipment setup that includes a specially designed access box known as the SD-Box. This box, along with a secure central infrastructure, forms a closed unit where users can remotely access the data via a dedicated project server, ensuring a seamless and secure data access experience. Meta plans to make Instagram data available in the CASD cleanroom as CSV files, which can be read and manipulated by standard statistical software (e.g. Excel, Stata, SPSS, R, Python).

Access for pilot program researchers

Researchers with an approved Stage 1 Registered Report will have access to Instagram data (a) that has been approved by Meta in response to the Data Request Form and (b) that study participants have consented and authorized to share for their account(s). As part of the pilot program, the Registered Report researchers will have access to such Instagram data in the third-party cleanroom designated by Meta for up to 15 months after the initial data delivery with extensions for access considered on a case-by-case basis. More details regarding data preservation if required by study funders or academic journals will be provided when Meta announces the third-party cleanroom at a later date.

Access for purposes of reproduction

After a Stage 2 Registered Report is published, the Instagram data will be made available only through restricted access for other researchers who have ethical approval from their IRB or equivalent to reproduce the findings.[6] Researchers must also make all other study data available to qualified academic researchers in the third-party cleanroom to reproduce their analyses. If researchers are unable to make the data they collected (e.g. survey data) openly available, researchers must specify how reproduction researchers may attain restricted access.

Reproduction requests will be approved by the Center for Open Science and will be limited to three per published Stage 2 Registered Report, which must be completed within three years of the date on which the Stage 2 Registered Report is published. The pilot program limits reproductions to three per Registered Report to allow for a “tie break” in the event of disagreements between the first two reproductions. Reproduction researchers will have access to the Instagram data in the third-party cleanroom for up to 6 months after approval within the three year period after the Stage 2 Registered Report is published.

[6] Reproduction refers to obtaining the same results for the same research questions and hypotheses using the same input data, methodological procedures and computations steps, and conditions of analysis (National Academies of Sciences, Engineering, and Medicine, 2019).

Data overview

The list of variables provided in this User Guide captures behaviors and account settings across the following categories:

1. **Account information** such as Instagram join date and connections (e.g. number of followers).
2. **Production of content** such as timestamps for what types of content users posted (e.g. whether they posted a Reel or Story).
3. **Consumption of content** such as session start and end times and content views.
4. **Interactions with content or other users** such as timestamps for comments, likes, and direct messages.
5. **Use of safety and control features** such as quiet mode and counts of blocked or restricted accounts.

Certain data types are not eligible for sharing as part of this pilot program, including but not limited to identifiers, such as profile or account names, demographic information, user generated content (e.g. textual or visual content of posts, comments, or messages), ad content, or any inferred metrics.

Researchers may request that a study participant link their Instagram data if the participant is over 13 years old and in one of the following 24 countries: Argentina, Australia, Brazil, Canada, Chile, Colombia, Egypt, Germany, France, India, Indonesia, Italy, Japan, Mexico, Nigeria, Peru, the Philippines, Poland, South Africa, Spain, Turkey, the United Kingdom, the United States, or Vietnam.

Researchers may request *either* up to 30 days of retrospective data *or* up to 30 days of prospective data. That is, up to 30 days of data from before *or* after the date on which each study participant authorized Meta to share data from their account(s) using the process outlined in the "Collecting and verifying consent from study participants" section. For longer studies, researchers must recontact participants to obtain consent and authorization for sharing additional period(s) of either retrospective or prospective Instagram data. For example, studies requiring 60 days of prospective Instagram data for each participant must recontact each participant for consent and authorization for an additional 30 days of data on the date that is 30 days after their first instance of consent and authorization.

See the "List of tables and variables" section for additional details.

Data disclosures

The variables described in this User Guide will be created using internal Meta company data sources available during the study period. Data governance and data sources described in this document may change from time to time and the data or variables presented here may not be representative of Meta's current operations. Due to data collection issues, such as logging errors in the data sources from which these variables are created or differences in user experience across app versions, these variables might suffer from fluctuating quality, including completeness and accuracy. Within this User Guide, we disclose known sources of fluctuation in the "List of tables and variables" section.

Researchers using these data are responsible for conducting standard and thorough data cleaning processes and are responsible for ensuring that their analyses are accurate. Researchers might find issues with the data when conducting their analysis. We encourage researchers to share issues with Meta or COS. These may include, but are not necessarily limited to, data quality, validity, or fidelity issues. If data quality issues are identified, please contact Meta or COS within 30 days of data delivery. Meta might not be able to address the issues identified, particularly in the case of historical data and lapsed data retention periods, but in such cases, Meta will work to disclose such issues in the "Updates and errata" section of this User Guide and/or in the codebook provided upon data delivery.

Event data availability

Some study participant accounts might not have event-level data during one or more days in their authorization period. Reasons for this include but are not limited to the account not having Instagram sessions on a particular day (i.e. not opening the app/site) or the account being deactivated.

Data source limitations

Data described in this User Guide comes from one of two sources: client-side logging and snapshots of Meta's production database. These sources have the following limitations:

1. **Client-side logging:** Certain data about how users interact with Meta's apps are captured by the client device (i.e. mobile device or computer) and sent back to Meta for logging. With all client-side logging there is potential for errors introduced by the interface itself (e.g. a new iOS or Android operating system release could have a bug that causes inaccurate logging) or by data loss (e.g. failure

during transmission of event data from client side to Meta's servers). Such errors are difficult or impossible for Meta to correct.

2. **Production snapshots:** Certain data described in this User Guide that are visible on Instagram, such as data about posts and account settings, are stored in Instagram's production database. For this pilot, such data will be retrieved as 'snapshots' of the database and will reflect the live state of Instagram at the time of the snapshot. Participants' data may fluctuate in between snapshots.

Data disclaimers

Before sharing data for research purposes, Meta works diligently and utilizes a variety of quality assurance measures to improve data quality, including its validity, accuracy, and reliability. However, given the volume of data released and the imperfection of any quality assurance process, it is possible that inaccuracies may 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 these data, individuals acknowledge that the data may contain some nonconformities, defects, or errors. Meta does not warrant that the data will meet all needs or expectations, that its use will be uninterrupted, or that nonconformities, defects, or errors can or will be corrected.

Data processing

Dates and times

Dates will be provided in Pacific Time and timestamps will be provided in Unix time (i.e. the number of non-leap seconds that have elapsed since 00:00:00 UTC on January 1, 1970).

Distinct events

With the exception of direct message sends[7], each row in event-level tables will represent a distinct event—even if the rows appear to be identical. This apparent duplication can occur if a participant account takes the same action in rapid succession. For example, a participant account might like two single-image Feed posts one after another before a full

[7] In the direct message sends table, 'duplicate' rows may also represent a single message sent to multiple accounts within a group chat. See the description of the direct message sends table for more details.

second has elapsed. In this case, both 'likes' would have the same value for event_time, as well as for other columns in the table.

Time periods

Data collection period

Meta will deliver consented and authorized Instagram data for study participants at the end of each "**data collection period**." For longitudinal studies with multiple waves of data collection, this means Meta will deliver the consented and authorized Instagram data for study participants at the end of each wave's data collection period. Researchers must notify Meta of the end of their data collection period by sending the relevant consent certification file(s).

Study period

For longitudinal studies, the "**study period**" will begin on the first day of the first data collection period and end on the last day of the last data collection period. For cross-sectional studies, the data collection period and study period will be the same.

Participant authorization period

Researchers may request *either* up to 30 days of retrospective data *or* up to 30 days of prospective data for each instance of consent to data sharing (using the applicable consent(s) provided by Meta) and authorization. Researchers should adhere to the privacy principle of data minimization and only request the number of days of retrospective or prospective data required for their proposed study. The period of retrospective or prospective data for each instance of consent and authorization is referred to in this document as the "**authorization period**."

The authorization period is not inclusive of the day of authorization and is instead calculated using the date before (for retrospective data) or after (for prospective data) the date the first account was authorized by a given study participant.

As described in the "List of tables and variables," certain data will be retrieved on the last day of the authorization period (e.g. "Account creation").

Identifiers

As noted above, Meta plans to make Instagram data available to researchers in the third-party cleanroom as CSV files. Each CSV file will have two study-specific identifiers:

- **Participant ID:** This is the unique identification number for each participant in the study generated by the researcher and provided to Meta using the redirect link to the Instagram authorization flow.
- **Authorized Account ID:** This is an identification number that Meta will generate to distinguish between multiple accounts authorized by the same study participant. The identification number will correspond to the order in which eligible accounts were authorized, starting with 1 (e.g. 1, 2, and so forth). For longitudinal studies, this identification number will remain consistent across survey waves; if additional accounts are authorized after the first wave, ID assignment will resume with the next consecutive integer.

Data merge and quality checks

Meta aims to deliver data to the third-party cleanroom approximately 45 days after the last day of the data collection period after merging the consent certification CSV(s) with data logged by the Instagram authorization flow and conducting final data processing and file quality checks.

After receiving researchers' consent certification file(s) described in "Step 2. Consent verification," Meta will merge the CSV(s) for each study with the data logged by Meta's Instagram authorization flow for that study. Meta will **drop** the following after completing the merge in the following order:

1. Any rows of data for participant IDs that are not found in the CSV file(s) provided by the researcher.
2. Any 13-17-year-old participant who does not have a timestamp for both their parent's or guardian's consent and their own consent in the CSV file(s) provided by the researcher.
3. Any 13-17-year-old participant who does not have a parent's or guardian's consent timestamp that is after their own consent timestamp.
4. Any participant 18 years or older who does not have a timestamp for their consent in the CSV file(s) provided by the researcher.
5. Any rows of data for authorized account IDs that are not eligible for data sharing because they are determined by Meta to be in a sanctioned or other sensitive market.[8]

[8] Researchers must only recruit study participants who reside in the eligible countries listed in the RFP and the "Data overview" section of this User Guide. For more information about how Instagram determines the primary location of any account, see the Instagram Help Center.

6. Any rows of data for subsequent accounts authorized after the first eligible account was authorized that are not linked to the first account in the Meta Accounts Center.
7. Any rows of data for authorized accounts that have been deleted prior to data delivery.[9]
8. Any rows of data pertaining to any authorized accounts' content (e.g. posts) or interactions (e.g. comments, likes, and messages) that have been deleted prior to data delivery.[7]
9. Any participants who have withdrawn their consent to share Instagram data or delete such data, either in whole or in part.

[9] Meta will specify the exact date on which the data were dropped in the codebook for each Registered Report study, which Meta will provide to researchers at the time of data delivery.

List of tables and variables

The following tables and variables have been designated as eligible for data sharing for this pilot program.

Account creation

Filename	account_creation.csv
Level	Participant account
Data	Date on which participants' Instagram accounts were created

Overview

This table contains the date(s) on which study participants created their Instagram account(s). Each row corresponds to a participant's account creation date; there is exactly one row per participant per account.

Disclosures

Data source

Account creation date is retrieved from Meta's production database on the last day of the participant's authorization period. However, account creation date is not expected to vary over time and can be considered to be valid for any date in a participant's authorization period.

Variables

Name	Label	Type	Description
date	Date	date	Date (Pacific Time) on which participant's account creation date was retrieved. *Format: YYYY-MM-DD*
participant_id	Participant ID	string	Study-specific participant ID.
authorized_account_id	Authorized account ID	string	Study-specific authorized account ID.
account_creation_date	Account creation date	date	Date (Pacific Time) on which study participant's account was created. *Format: YYYY-MM-DD*

Account settings

Filename	account_settings.csv
Level	Participant account
Data	Daily snapshots of participants' Instagram account settings

Overview

This table contains daily indicators for quiet mode, account privacy, and stories archive settings, described below. Each row corresponds to the settings for a given participant account on a given day, with one row per account for each date in the participants' authorization period.

Quiet mode

Quiet mode lets users automatically pause their Instagram notifications every day during the time period they choose. The default setting is every night from 11 p.m. to 7 a.m., but these hours can be edited in Settings. When users are in quiet mode, they can still use Instagram and send and receive messages. Their activity status will change to "In quiet mode," and the green active dot will change to a moon icon. People who message users in quiet mode will get an auto-reply. When quiet mode ends, the users get a summary of any new notifications or messages they received while they were in quiet mode.

Account privacy

Users can choose between a public and private account in Instagram Settings. See the Instagram Help Center for a detailed description of the differences between public and private accounts on Instagram.

Disclosures

Data source

Account settings are retrieved daily via snapshots of Meta's production database.

Variables

Name	Label	Type	Description
date	Date	date	Date (Pacific Time) on which participant's account settings were retrieved. *Format: YYYY-MM-DD*
participant_id	Participant ID	string	Study-specific participant ID.
authorized_account_id	Account ID	string	Study-specific authorized account ID.
account_is_quietmode_enabled	Quiet mode	boolean	Flag indicating whether quiet mode was enabled on the account. *Values:* - *0: quiet mode was not enabled* - *1: quiet mode was enabled*
account_is_private	Account privacy	boolean	Flag indicating whether the account was private. *Values:* - *0: account was not private* - *1: account was private*
account_is_story_archive_enabled	Story archive	boolean	Flag indicating whether automatic story archiving

Name	Label	Type	Description
			was enabled. *Values:* - *0: automatic story archiving was not enabled* - *1: automatic story archiving was enabled*

Foreground sessions

Filename	foreground_sessions.csv
Level	Instagram foreground session
Data	Start and end times of participants' "foreground sessions" on Instagram

Overview

This table contains the start and end times of participant accounts' "foreground sessions" on Instagram. For the purposes of this study, "foreground sessions" are defined as starting when the Instagram app/site application is "foregrounded" for a participant account and ending when the app/site is "backgrounded", is closed, crashes, or times out (for interfaces without background events, i.e. web) for that account.

Each row corresponds to a single "foreground session" for a given participant account. Participant accounts can have zero or multiple sessions on a given day. Session length (session_end_time - session_start_time) can be summed over a period of interest (e.g. day) to represent the amount of time a participant account spent on Instagram.

Disclosures

Data source

This table is generated from client-side logging of foregrounding and backgrounding events for the Instagram site/app.

Data coverage

This table only includes sessions from the following interfaces: native Instagram apps (iOS and Android), desktop website, mobile website, and Instagram Lite.

Session truncation

Sessions beginning before or ending after the bounds of a participants' authorization period are truncated such that this table only includes the portion of the session contained within the participant authorization period. For example, assume a participant account has a 15-minute session on the last day of their authorization period, foregrounding Instagram at 23:59:00 and backgrounding the app at 00:14:00 (the next day). In this table, that session will appear as approximately one minute in length, with start time at 23:59:00 and end time at 23:59:59.

Variables

Name	Label	Type	Description
participant_id	Participant ID	string	Study-specific participant ID.
authorized_account_id	Account ID	string	Study-specific authorized account ID.
session_start_time	Session start time	timestamp	Time (Unix time) at which the Instagram app or site was "foregrounded" for the participant account.
session_end_time	Session end time	timestamp	Time (Unix time) at which the Instagram app or site was closed or "backgrounded", the app crashed, or there was a timeout on an interface without background events (i.e. web) for the participant account.

Organic content views

Filename	organic_content_views.csv
Level	Organic content view event
Data	Timestamps for participants' organic content views on Instagram

Overview

This table contains timestamps for logged views of organic content (non-ads) by participant accounts. Organic content view events are logged when a piece of content is either: (a) fully visible to the user, or (b) covers at least 50% of the user's screen for at least 250 milliseconds. Each row corresponds to a participant account's view of a Feed post (single media), Carousel post (album), Story, Reel, Live broadcast, or IGTV post.

Disclosures

Data source

This table is generated from client-side logging of organic content view events.

Self-views

Participant accounts' views of their own content (i.e. content they have authored) are not counted as views of posts and are therefore excluded from this data by default.

Connections

Logging of indicators for whether the participant account is following or followed by the content author may significantly lag logging of view time. These connection indicators reflect the connection status as of the time at which connection logging occurred, rather than when the view itself was logged.

Media count

Media count describes the number of images or videos contained in a carousel post. It does not indicate the number of media viewed by the participant account; that is, the participant account may have viewed one, multiple, or all the images or videos in the carousel.

Warning screens

Views of organic content covered by a warning screen might be logged differently depending on interface (e.g. iOS, Android) and content type (e.g. Reels, Stories). For example, in the case of sensitivity overlays:

- **On iOS**, organic content views of **Feed and Carousel posts** and fullscreen **Reels** are not logged when a sensitivity overlay is covering the content (i.e., views of organic content covered by a sensitivity overlay are only logged if a user clicks the overlay to uncover the content).
- **On Android**, organic content views of **Feed and Carousel posts** and fullscreen **Reels** are logged when a sensitivity overlay is covering the content (i.e., views of organic content covered by a sensitivity overlay are logged regardless of whether a user clicks the overlay to uncover the content).
- **On both iOS and Android**, organic content views of **Stories** are logged when a sensitivity overlay is covering the content (i.e., views of organic content covered by a sensitivity overlay are logged regardless of whether a user clicks the overlay to uncover the Story).

Variables

Name	Label	Type	Description
participant_id	Participant account	string	Study-specific participant ID.
authorized_account_id	Account ID	string	Study-specific authorized account ID.
event_time	View time	timestamp	Time (Unix time) at which the organic content view was logged.
interface	Interface	string	Interface on which the organic content view was logged. *Values:* - *instagram_ios: native iOS Instagram app* - *instagram_android: native Android Instagram app* - *instagram_www: desktop website* - *instagram_msite: mobile website* - *instagram_lite: Instagram Lite app* - *other: other interface (e.g. IGTV app)*
content_type	Content type	string	Type of viewed content (e.g. Feed post, Story). *Values: feed, carousel, story, reel, live, igtv*
content_media_type	Content media type	string	Whether the viewed content was an image, video, album, or broadcast. *Values: image, video, album, broadcast*
content_media_count	Content media count	integer	Count of individual media contained in the viewed content. For carousel posts, this is the number of

Name	Label	Type	Description
			individual images and/or videos contained in the album. For all other content types, this number is 1 (i.e. single media). *Values:* - *1: content type was not 'carousel'* - *2-20: content type was 'carousel'*
content_audience_is_close_friends	Content audience	boolean	Flag indicating whether the content was shared with the author's close friends list. *Values:* - *0: content was shared with other audience* - *1: content was shared with close friends list*
content_is_reshare	Content is a reshare	boolean	Flag indicating whether the content was a reshare (not original content). *Values:* - *0: content was original (not a reshare)* - *1: content was a reshare*
participant_is_following_author	Participant account is following the content author	boolean	Flag indicating whether the participant account was following the author of the viewed content at the time of logging. *Values:* - *0: participant account was not following content author* - *1: participant account was following content author*

Name	Label	Type	Description
participant_is_followed_by_author	Participant account is following by the content author	boolean	Flag indicating whether the participant account was followed by the author of the viewed content at the time of logging. *Values:* - *0: participant account was not followed by author of viewed content* - *1: participant account was followed by author of viewed content*

Billable ad impressions

Filename	billable_ad_impressions.csv
Level	Ad impression event
Data	Timestamps for participants' billable ad impressions on Instagram

Overview

This table contains timestamps for participant accounts' billable ad impressions on Instagram during the study period. Ad impressions are counted when any part of the ad appears on the user's screen (i.e. greater than zero pixels for greater than zero seconds).

Each row corresponds to a billable ad impression for a given participant account. Participant accounts can have zero or multiple billable ad impressions per day.

Disclosures

Data source

This table is generated from client-side logging of ad impression events. In addition to the general limitations to client-side logging described earlier, Meta is also aware of the following limitations specific to ad impression logging, but believes these to be insignificant.[10]

[10] See Instagram Description of Methodology in the Meta Business Help Center.

- Caching: Instagram employs appropriate cache-busting techniques; however, defeating all caching is not possible.
- Notification tray: On Android devices, when the user drags the notification tray into view and obstructs the Feed and Stories, Meta will keep logging duration until the user scrolls away.
- Force close: If a user force closes the application or browser too quickly after generating an ad impression, the server-client connection is severed and the impression event may not be captured by backend loggers. This is a small edge case.

Data coverage

Some study participant accounts might not have ad impressions during some or all of their Instagram sessions, due to one of the following reasons:

1. Participant's account is ineligible to see ads.
2. Participant's account has opted out of sharing data for ad targeting.

If, during the study period, a participant account opts out or becomes ineligible to see ads during the data collection period, their ad impressions data will be excluded entirely from this table (i.e. both before and after opt-out date).

This table will not include data for accounts located in the EU because teens in the EU are ineligible to see ads.[11]

Billable impressions

After an ad impression is recorded in Meta's logs, Meta flags some impression events as non-billable based on a list of rules; the remaining impression events are considered billable. Examples of rules used to identify non-billable events include:

[11] See About Advertising to Teens in the Meta Business Help Center.

1. Invalid impressions: Instagram has a number of systems in place to detect and remove General Invalid Traffic; e.g. impressions generated by bots or spiders, etc., as well as auto checks for adware, malware, spam accounts, fake accounts, etc.
2. Activity-based filtration: If incoming events from a single user source exceed a defined threshold within a given amount of time, subsequent events will be logged as non-billable and will not be reported.

See Instagram Description of Methodology on the Meta Business Help Center for additional details.

Variables

Name	Label	Type	Description
participant_id	Participant account	string	Study-specific participant ID.
authorized_account_id	Account ID	string	Study-specific authorized account ID.
event_time	Impression time	timestamp	Time (Unix time) at which the ad impression was logged.

Push notification events

Filename	push_notification_events.csv
Level	Push notification event
Data	Timestamps for participants' Instagram notification events

Overview

This table contains timestamps for participant accounts' push notification events for Instagram, including when notifications are eligible to be displayed by a user's device, and when a user clicks on a notification.

Each row corresponds to an Instagram push notification event for a given participant account. Participant accounts can have zero or multiple notification events per day.

Disclosures

Data source

This table is generated from client-side logging of push notification events. Instagram notifications are delivered via third party libraries, which typically offer best-effort delivery guarantees but can pose challenges to reliability for both delivery and logging.

Data coverage

Some participant accounts might not have push notifications during some or all of their Instagram sessions if notifications have been turned off, such as via Instagram account settings or a device operating system. This table does not include data for notifications from Instagram Direct, including messages or calls.

Event types

This table includes two types of push notification events:

- Display eligible: Event triggered when a push notification passed all client-side checks; indicates the device received the notification and the notification was eligible to be displayed by the device.
- Clicked: Event triggered when a user clicked on the push notification.

Instagram Direct notifications

This table does not include data for notifications from Instagram Direct, including messages or calls.

Variables

Name	Label	Type	Description
participant_id	Participant ID	string	Study-specific participant ID.
authorized_account_id	Account ID	string	Study-specific authorized account ID.
event_time	Notification time	timestamp	Time (Unix time) at which the notification event was logged.
notification_type	Notification type	string	Type of notification event.

Name	Label	Type	Description
			Values: - *display_eligible: notification eligible to be displayed by device* - *clicked: notification clicked*

Organic content creation events

Filename	content_creation_events.csv
Level	Content creation event
Data	Timestamps for participants' content creation

Overview

This table contains timestamps for when participant accounts created organic content for Instagram, including Feed posts (single media), Carousel posts (album), Stories, Reels, or IGTV posts.

Each row corresponds to the creation of a piece of content by a given participant account. Participant accounts can have zero or multiple content creation events per day.

Disclosures

Data source

Data about content creation are retrieved from Meta's production database on the last day of the participant's authorization period. This table reflects a snapshot of data about content created during the participant's authorization period that has not been deleted at the time of retrieval.

Event time

Event time represents the time at which the participant account began creating the content (i.e. create time). Create time might differ from the time the content was published on Instagram. For example, a user might take a few minutes to edit the content before publishing or schedule content to be published in the future.

Stories creation

Data about stories creation are included in this table only for stories that have been archived. Automatic story archiving can be turned on or off by users in account settings. Users can also manually archive a story while it is still active, even if the automatic story archiving function is disabled.

Variables

Name	Label	Type	Description
participant_id	Participant ID	string	Study-specific participant ID.
authorized_account_id	Account ID	string	Study-specific authorized account ID.
event_time	Creation time	timestamp	Time (Unix time) at which the content was created.
content_type	Content type	string	Type of content (e.g. Feed post, Story). *Values: feed, carousel, story, reel, igtv*
content_media_type	Content media type	string	Whether the content was an image, video, or album.

Name	Label	Type	Description
			Values: image, video, album
content_media_count	Content media count	integer	Count of individual media contained in the content. For carousel posts, this is the number of individual images and/or videos contained in the album. For all other content types, this number is 1 (i.e. single media). *Values:* - *1: content type was not 'carousel'* - *2-20: content type was 'carousel'*
content_audience_is_close_friends	Content audience	boolean	Flag indicating whether the content was shared with the participant account's close friends list. *Values:* - *0: content was shared with other audience* - *1: content was shared with close friends list*
content_is_reshare	Content is reshare	boolean	Flag indicating whether content was a reshare (not original content). *Values:* - *0: content was original (not a reshare)* - *1: content was a reshare*

Comment events

Filename	comment_events.csv
Level	Comment event
Data	Timestamps for comments made by participants or on participants' content

Overview

This table contains timestamps for when participant accounts commented on content (authored by the participant account or another user) and for when participant accounts' content was commented on by another user.

Each row corresponds to the creation of a comment by a given participant account or on the content from a participant account. Participant accounts can have zero or multiple comment creation events per day.

Disclosures

Data source

Data about comments are retrieved from Meta's production database on the last day of a participant's authorization period; this table reflects a snapshot of Instagram comments at the time of retrieval.

Event time

Event time represents the time at which the participant account began creating the comment (i.e. create time). Create time might differ from the time the comment was published on Instagram. For example, a user might take a few minutes to edit the comment before publishing.

Variables

Name	Label	Type	Description
participant_id	Participant ID	string	Study-specific participant ID.
authorized_account_id	Account ID	string	Study-specific authorized account ID.
event_time	Comment time	timestamp	Time (Unix time) at which the comment was created.
content_type	Content type	string	Type of content on which comment was made (e.g. Feed post, Story). *Values: feed, carousel, story, reel, igtv*
content_media_type	Content media type	string	Whether the content on which the comment was made is an image, video, or album. *Values: image, video, album*
comment_direction	Comment direction	string	Flag indicating whether the participant account sent or received the comment, or whether the participant account commented on their own content.

Name	Label	Type	Description
			Values: sent, received, self
comment_is_reply	Comment is reply	boolean	Flag indicating if the comment was a reply to another comment. *Values:* - *0: content was not a reply to another comment* - *1: comment was a reply to another comment*

Like events

Filename	like_events.csv
Level	Like event
Data	Timestamps for likes made by participants or on participants' content

Overview

This table contains timestamps for when participant accounts liked content (authored by the participant account or another user) and for when participant accounts' content was liked by another user.

Each row corresponds to a like *by* a given participant account or *on* the participant account's content. Participant accounts can have zero or multiple like events per day.

Disclosures

Data source

Data about likes are retrieved from Meta's production database on the last day of a participant's authorization period; this table reflects a snapshot of Instagram likes at the time of retrieval.

Variables

Name	Label	Type	Description
participant_id	Participant ID	string	Study-specific participant ID.
authorized_account_id	Account ID	string	Study-specific authorized account ID.
event_time	Like time	timestamp	Time (Unix time) at which the like was made.
content_type	Content type	string	Type of content on which like was made (e.g. Feed post, Story). *Values: feed, carousel, story, reel, igtv, comment*
content_media_type	Content media type	string	Whether the content on which the like was made was an image, video, or album. *Values: image, video, album, comment*
like_direction	Like direction	string	Flag indicating whether the participant account sent or received the like, or whether the participant account liked their own content. *Values: sent, received, self*

Direct message sends

Filename	direct_message_sends.csv
Level	Direct message send event
Data	Timestamps for direct messages sent by or sent to participants

Overview

This table contains timestamps for messages sent by and sent to participant accounts, as well as indicators for the type of connection between sender and receiver. Each row corresponds to a message send; participant accounts can have zero or multiple message events per day.

Disclosures

Data source

This table is generated from client-side logging of messaging events.

Data coverage

Some messages included in this table may never reach a user's Inbox or Message Requests folder. For example, some users may choose not to receive message requests at all.

This table excludes data about messages subject to end-to-end encryption or sent in vanish mode (i.e. ephemeral messages).

Connections

Logging of indicators for whether the participant is following or followed by the content author may significantly lag logging of message sends. These connection indicators reflect the connection status as of the time at which connection logging occurred, rather than when the message itself was sent.

Messages sent to multiple recipients

Messages can be sent to multiple recipients in three ways:

1. By using the share button to share content (such as a Feed post or Story) with multiple accounts simultaneously via individual messages (i.e. messages sent to multiple separate chats via a single share); or
2. By using the share button to share content with a group chat (i.e. as single message sent to a single group chat); or
3. By sending text or other content directly within a group chat (i.e. a single message sent to a single group chat).

In all cases, messages are logged separately for each account that received the message. Therefore, if a participant account sends a message to or shares content with a group chat, this table will include multiple rows for the same message. Similarly, if a participant account shares content with multiple accounts at the same time, this table will include multiple rows for the same share. Depending on the relationships between the participant account and the recipient accounts, these rows might be identical or they might differ within the sender_follows_receiver and receiver_follows_sender variables defined below. For example, if a participant account sends a message to a group chat with three other members, this table will contain three rows with the same value for event_time, message_type, and message_direction. Values for sender_follows_receiver and receiver_follows_sender will depend on whether the participant account is following the group chat member and vice versa.

Variables

Name	Label	Type	Description
participant_id	Participant ID	string	Study-specific participant ID.
authorized_account_id	Account ID	string	Study-specific authorized account ID.
event_time	Message send time	timestamp	Time (Unix time) at which the message was sent.
message_type	Message type	string	Type of message (e.g. text, reaction). *Values: text, story_reply, link, reaction, video, photo, sticker, gif, share, voice_message*
message_direction	Message direction	string	Whether the participant account sent or received the message. *Values: sent, received*
sender_follows_receiver	Sender follows receiver	boolean	Flag indicating whether the sender followed the receiver at the time the message was sent. *Values:* - *0: sender didn't follow receiver* - *1: sender did follow receiver*
receiver_follows_sender	Receiver follows sender	boolean	Flag indicating whether the receiver followed the sender at the time the message was sent. *Values:* - *0: receiver didn't follow sender*

Name	Label	Type	Description
			- *1: receiver did follow sender*

Connection counts

Filename	connection_counts.csv
Level	Participant account
Data	Daily

Overview

This table contains daily counts related to participants' connections, including followers, following, close friends, blocked accounts, and restricted accounts.

Disclosures

Data source

Connection counts are retrieved daily from Meta's production database; this table reflects a daily snapshot of participants' connection counts.

Variables

Name	Label	Type	Description
date	Date	date	Date on which connection counts were retrieved.

Name	Label	Type	Description
participant_id	Participant ID	string	Study-specific participant ID.
authorized_account_id	Account ID	string	Study-specific authorized account ID.
followers_count	Followers count	integer	Count of accounts the participant account was being followed by.
following_count	Following count	integer	Count of accounts the participant account was following.
close_friends_count	Close friends count	integer	Count of accounts on the participant account's close friends list.
blocked_count	Blocked accounts count	integer	Count of accounts the participant account had blocked.
restricted_count	Restricted accounts count	integer	Count of accounts the participant account had restricted.

Updates and errata

Any future changes to this User Guide will be described in this section and a new version will be released with a new version number. You can sign up for email alerts from COS here.

1. Version 1.1 has the following updates or corrections:
    a. In the "Data disclosures" section, we added the following sentence: "Data sources are maintained for business purposes and might have limitations for research purposes."
    b. For the "Organic content views" table, we added a new disclosure about how "Warning screens" may affect the logging of organic content views on different interfaces. We also added a variable for the interface on which the organic content view was logged.
    c. For the "Organic content views" and "Organic content creation" tables, we updated the maximum number of individual images and/or videos that can be contained in a carousel post from 10 to 20. We also updated the carousel glossary entry.
    d. We updated the "Data coverage" limitation for the "Billable ads impressions" table.
2. Version 1.2 has the following correction:
    a. In the "Organic content views" table, we corrected the description for "content_audience_is_close_friends," which incorrectly read, "Flag indicating whether the content was shared with the ***participant's*** close friends list." The description now reads, "Flag indicating whether the content was shared with the ***author's*** close friends list."
3. Version 2.0 has the following updates or corrections:
    a. We redrafted the "Collecting and varying consent from study participants" section and renamed it to the "Collecting and varying consent and authorization from study participants." The new section reflects the content of the webinar for pilot researchers hosted by COS on February 26, 2025. In particular, the new section introduces two additional Instagram account authorization flows that researchers can use for their studies, as well as a "test mode" that researchers can use to test the programming of their data collection instrument.
    b. In the "Accessing Instagram data" section, we named CASD as the secure cleanroom environment that will host the data, which was previously announced on the COS website.

c. In the “Data overview” section, we clarified that, similar to user generated content, “ad content” is not available as part of this pilot program.
d. For the “Like events” table, we added “comment” as a value for the “content_type” and “content_media_type” variables.
e. For the “Direct message sends” table, we added an explanation about how messages or shares to multiple accounts sent by the participant account will appear in the table.

Glossary

Term	Definition[12]
Authorized account ID	A study-specific identification (ID) number generated by Meta to distinguish between multiple accounts authorized by the same study participant. The identification number corresponds to the order in which eligible accounts were authorized, starting with 1 (e.g. 1, 2, and so forth). For longitudinal studies, this identification number remains consistent across survey waves; if additional accounts are authorized after the first wave, ID assignment resumes with the next consecutive integer.
Authorized Instagram data	Data approved for sharing with a Registered Report study from Instagram accounts authorized by study participants who provided: (1) active, valid consent or explicit agreement to participate in the study and (2) active, valid consent for Meta to share their data for the purposes of the study.
Accounts Center	A place where users can manage connected experiences (like logging in across their accounts) and change account settings for their Facebook, Instagram and Meta accounts. Users can add Facebook, Instagram and Meta accounts to the same Accounts Center. See Meta Help for more details.
Backgrounding	Occurs when a user loses the ability to interact with the Instagram app. Backgrounding may happen for multiple reasons, including user-initiated activity (e.g. switching apps) or device-related (e.g. screen lock).
Block	To disable another account from seeing or interacting with the user's account. See Instagram Help Center for more details.
Broadcast	Media type associated with Instagram Live content. Users can start a live broadcast to connect with their followers in real time. See Instagram Help Center for more details.
Cache-busting	Cache-busting is a technique used to prevent a web browser from re-using an ad it has already shown to someone and forces the browser to request a new ad each time the page refreshes.
Carousel	Instagram content type containing between 2 and 20 images and/or videos as a single post in Instagram Feed. Carousel posts have media type 'album'. See Instagram Help Center for more details.

[12] For more details about Instagram settings, features, or tools, researchers should consult the Instagram Help Center.

Term	Definition[12]
Cleanroom	A secure environment where researchers can analyze data with restrictions on data uploads and downloads.
Client-side logging	A method for event logging where user interactions with Meta's apps are recorded by the client application (on a mobile device or desktop computer) and transmitted back to Meta. Examples of data logged on the client side include impressions and notifications.
Close friends	Users can create a Close Friends list on Instagram so that when they share things like a post, reel, or story, they can choose for only people on that list to see it. See Instagram Help Center for more details.
Content author	The Instagram account that shared the content on the platform. Content author can be either the account that created the content (in the case of an original post) or the account that reshared the content (in the case of a reshare).
Event data	Timestamped information collected and recorded when a specific event or action occurs on Instagram
Feed post	Instagram content type for single media (image or video) posts shared to a user's Instagram Feed. Feed posts have media type 'image' or 'video'. See Instagram Help Center for more details.
Followers	The Instagram accounts that are following a user's account. A user's followers have opted in to seeing content from that user. See Instagram Help Center for more details.
Following	The Instagram accounts that are followed by a user's account. A user has opted in to seeing content from the accounts they are following. See Instagram Help Center for more details.
Foregrounding	Occurs when an app becomes fully visible on a user's device, allowing them to interact with it and use its features.
Foreground session	Time period starting when the Instagram app/site is "foregrounded" and ending when the app/site is "backgrounded", is closed, crashes, or times out (for interfaces without background events, i.e. web).
IGTV	Instagram content type for long-form, vertical videos up to 60 minutes in length. Although Meta stopped supporting its standalone IGTV app in 2022, some IGTV content may still be available in the Instagram app. IGTV content has media type 'video'.
Instagram account authorization	For this pilot program, this refers to the process of obtaining permission from a user to access the Instagram data associated with the Instagram account they are currently logged into. Permission is obtained on a page on the Instagram app, which study participants

Term	Definition[12]
	can navigate to using a redirect URL provided by Meta and included by researchers in their data collection instrument.
Institutional Review Board (IRB)	An administrative body established to protect the rights and welfare of human research subjects recruited to participate in research activities conducted under the auspices of the institution with which it is affiliated.
Like	An Instagram interaction on a piece of content or comment. To 'like' content or comments, users can double tap on the content or tap on the content or comment's heart icon. Users can 'like' posts, stories, and comments. See Instagram Help Center for more details.
Live	Instagram content type for live broadcasts shared with a user's followers in real time. Once their live broadcast ends, a user can share a replay of it to their Instagram profile and feed. Live content has media type 'broadcast'. See Instagram Help Center for more details.
Media type	The type (i.e. format) of media contained in a piece of content. For Instagram, this includes images, videos, albums, and broadcasts.
Organic content	Content posted by an Instagram user to share with their connections or the public, as distinguished from paid content such as ads.
Organic content view	A type of event logged when a piece of organic content is either: (a) fully visible to the user, or (b) covers at least 50% of the user's screen for at least 250 milliseconds.
Participant authorization period	For this pilot program, this refers to either a retrospective or prospective period of time between 1 and 30 days from the date on which a study participant authorized their first eligible account.
Participant ID	Study-specific participant identifier (ID) generated by a researcher for the purposes of their study.
Private account	Instagram account setting where only a user's followers can see their posts, followers, or following list. Only accounts approved by the user can follow that user's account if set to private. See Instagram Help Center for more details.
Production database	Where Meta stores and manages live data for the Instagram app/web experiences accessible to users.
Quiet mode	An Instagram feature that allows users to automatically pause their Instagram notifications every day during the time period they choose. See Instagram Help Center for more details.

Term	Definition[12]
Reel	Instagram content type for short-form videos up to 90 seconds in length. Reels have media type 'video'. See Instagram Help Center for more details.
Reproduction	Reproduction is a concept from the open science movement and refers to obtaining the same results for the same research questions and hypotheses using the same input data, methodological procedures and computations steps, and conditions of analysis.
Reshare	Instagram content originally created and shared elsewhere on the app either by the original content author or by a different Instagram account. Examples: (1) Instagram content originally shared as a feed post shared as a story; (2) a story shared by an account mentioned in the original content author's story.
Snapshot	For this pilot program, this refers to a one-time retrieval of data from Instagram's production database which reflects the live state of Instagram at the time of the snapshot.
Story	Instagram content type for sharing images and videos which disappear from a user's profile, Feed and messages after 24 hours. Stories have media type 'image' or 'video'. See Instagram Help Center for more details.
Unix time	A date and time representation commonly used in computing, defined as the number of non-leap seconds that have elapsed since January 1, 1970, at 00:00:00 UTC.