Meta Content Library

transparency.fb.com

Agenda

01 Product overview

02 How to gain access

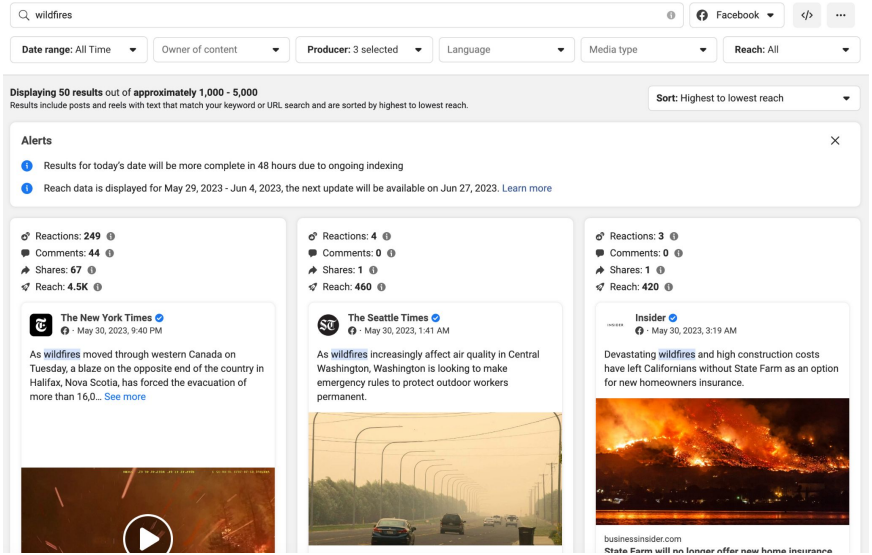
03 Product demonstrations

04 Q & A

The Products

Meta Content Library

Meta's comprehensive database of publicly available content across Facebook and Instagram, available through a web-based explorer and an API




```
# page data
# hashed id for NYT (obtained from CodeGen)
page_search <- client$search_pages(page_ids="AZsM9qZz_EcnfVGSMmHkOhxHFAYnLQPGMOxQJp9D-D7TZhu5J0_AqQ")
pages <- page_search$query_next_page('dataframe')
names(pages)
head(pages[, c("id", "name", "category_list")],1)
```

Last executed at 2023-06-30

'about' · 'category_list' · 'creation_date' · 'follower_count' · 'id' · 'name' · 'page_categories' · 'verification_status' · 'website' · 'page_transparency.confirmed_page_owner' · 'page_transparency.has_active_ads' · 'page_transparency.has_run_political_ads' · 'page_transparency.admin_countries.data' · 'picture.url'

A data.frame: 1 × 3

	id	name	category_list
	<chr>	<chr>	<list>
1	AZsM9qZz_EcnfVGSMmHkOhxHFAYnLQPGMOxQJp9D-D7TZhu5J0_AqQ	The New York Times	Media/News Company , News & media website

```
# posts data
post_search <- client$search_posts(page_ids="AZsM9qZz_EcnfVGSMmHkOhxHFAYnLQPGMOxQJp9D-D7TZhu5J0_AqQ")
posts = post_search$query_next_page('dataframe')
names(posts)
# show a few column values
head(posts[, c("id", "text", "creation_time", "statistics.reaction_count")],1)
```

Last executed at 2023-06-30

'id' · 'text' · 'creation_time' · 'modified_time' · 'lang' · 'statistics.like_count' · 'statistics.love_count' · 'statistics.wow_count' · 'statistics.haha_count' · 'statistics.sad_count' · 'statistics.angry_count' · 'statistics.care_count' · 'statistics.comment_count' · 'statistics.reaction_count' · 'statistics.share_count' · 'post_owner.data' · 'link_attachment.description' · 'link_attachment.link' · 'link_attachment.name' · 'link_attachment.caption'

A data.frame: 1 × 4

	id	text	creation_time	statistics.reaction_count
	<chr>	<chr>	<chr>	<int>
1	AZswvQU_lLXNXihLzXpbhm7c5UHQqHzJnBfncrrTgqy6A1yFxTVQYRuKoDsh	A California woman and Instagram influencer, who in 2020 falsely reported and posted on social media that a couple had attempted to kidnap her two young children, was sentenced on Thursday to 90 days in jail, prosecutors said.	2023-06-30T19:25:07+0000	39

Meta

Data sharing: principles into practice

Privacy and security

- Virtual data clean rooms
- 30-day user data deletion
- Targeted scrubbing personal identifiers and encryption of entity IDs
- Eligibility and access requirements

Independence

- Independent review of researcher applications
- No restrictions on research topic
- No publication review
- Applications accepted from global researchers*

Transparency

- Data and search quality disclosures
- New metrics (e.g. view count of posts)
- Participation in EDMO Code of Conduct data sharing pilot
- Signatory to the EU Code of Practice on Disinformation

*Certain exceptions may apply

How to gain access

Join the early access (beta) program

Beta Program Overview

- **60-day access to the tools for testing purposes (Nov-Dec 2023 only)**
- Documentation and support from our technical team if questions arise
- 45-minute feedback session and survey about your experience using the tools
- **MetaResearchApplications@meta.com**

Apply for general access with ICPSR at the University of Michigan

https://socialmediaarchive.org/

UNIVERSITY OF MICHIGAN

Eligibility Requirements

- Be an individual affiliated with a qualified academic or research institution
- Propose a scientific or public interest research topic
- Ensure the research is free from any commercial interests
- Disclose funding source(s)
- Fulfill data security and confidentiality requirements
- Demonstrate that data requested is relevant and limited to what is necessary for answering the research question(s)

Application Details

- Applications will be submitted to and vetted by research staff at ICPSR at the University of Michigan

Platforms for Data Analysis

Feature	Meta Researcher Platform (early access)	Third-Party Data Clean Room (general availability)
Platform/Computation	Browser-based computation on modified version of Jupyter with R, Python, Julia	Ubuntu 22; operates on virtual remote desktop and supports access to R, Python, Stata
Data Hosting	Hosted on Amazon Web Services; offers CPU and GPU servers; free computation	Free compute, offers CPU and GPU upon request
Data Coverage	No multimedia	Multimedia access
Export of Research Outputs	Research outputs (e.g. figures, graphs, tables, code, statistics) may be exported out of the secure environment following applicable data reviews	
Data Upload	Only Machine Learning models (i.e. code) permitted for upload, no datasets	Dataset upload on case by case basis, following data review, under consideration
Data Deletion	Meta will effectuate user data deletion on researchers' Jupyter notebooks by removing output cells and local files on the 1st of every month.	Meta will provide researchers a deletion endpoint to track deletion status of data in their environments.

Product demonstration

Content Library API Data and Entities

Meta's Content Library API is an API product that facilitates research on public Facebook and Instagram content. Users can analyze billions of near real-time and historical data points. Researchers can query data from:

POSTS

Public Posts on public Facebook Pages, Groups, or Events

PAGES

Public Pages

GROUPS

Public Groups

EVENTS

Public Events. They do not need to be hosted by a Page or public Group.

FACEBOOK

POSTS

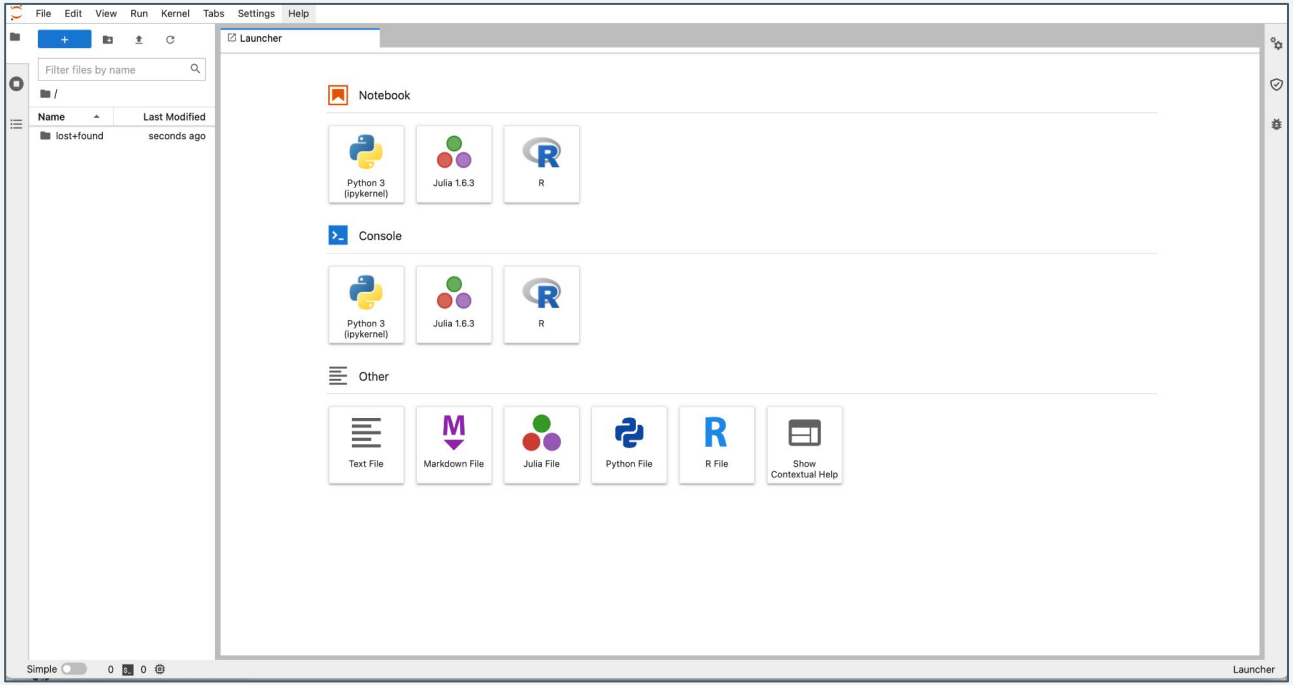
Public Posts from public Instagram creator and business accounts

ACCOUNT

Public creator and business accounts

INSTAGRAM

Researcher Platform Overview



A view of the JupyterLab instance on the Researcher Platform where researchers can analyze available data in Python or R, including custom Python and R libraries. Pre-installed libraries allow researchers to perform common processes such as data processing, data analysis, machine learning, and data visualization.

Researchers can use the Researcher Platform, a secure digital clean room, to access Content Library API, which contains certain Facebook and Instagram public data.

The Researcher Platform runs a modified version of Jupyter, an open source tool that **supports statistical packages such as Python and R**.

Return to Table of Contents

Endpoint Search Method

Each of the listed search method will return a Search object across both Facebook and Instagram endpoints.

Endpoint	Method	Description
Facebook Page	search_pages()	Defines a Page search query and returns a Generator that can execute the query.
Facebook Post	search_posts()	Defines a Post search query and returns a Generator that can execute the query.
Facebook Group	search_groups()	Defines a Group search query and returns a Generator that can execute the query.
Facebook Event	search_events()	Defines an Event search query and returns a Generator that can execute the query.
Instagram Account	search_ig_accounts()	Defines an Instagram Account search query and returns a Generator that can execute the query.
Instagram Post	search_ig_posts()	Defines an Instagram Post search query and returns a Generator that can execute the query.

For more information about query syntax, see Meta for Developers page.

Helpful Links

Product information: https://transparency.fb.com/

Product documentation and educational resources: https://developers.facebook.com/docs/content-library-and-api

Early access (beta) program contact information: openresearch@meta.com

Application information: https://socialmediaarchive.org/ (ICPSR)

Thank you!

Q&A

Researcher Platform: Hardware Specifications

Memory: 50G guaranteed; 64G limit

Storage:

capacity: 32GB (applies only to EBS. No current limit on S3 storage)

CPU: 16 cores; per core (m5.4xlarge)

GPU: G4dn.4xlarge, 64GiB memory; GPU memory: 16 GiB; Instance storage: 125 GB; Network performance: Up to 25 Gbps

R-API Analysis Process

Users will need to pull in data by first querying for specific keywords. After getting the API results as a dataframe, users can use Python or R language to analyze the results to derive meaningful insights.

STEP 1: Create a Notebook in the Researcher Platform by selecting "Python 3" notebook to use Python and "R" notebook to use R.

STEP 2: Query endpoint by searching for specific keywords. Example below returns the first 5 rows for the search term "2024 election" from FB Pages.

```
response = client.search_pages(
    q="2024 election",

fields="id,category,location,engagement{count}",
    limit=5,
    data_type="dataframe"
)
print("==== pages ====")
```

==== pages ====

	id	category	location.city	location.country	location.state	location.zip	engagement.count	location.latitude	location.longitude	location.street
0	7191206138	Nonprofit Organization	New York	United States	NY	10018	1766537	NaN	NaN	NaN
1	108509679181159	Nonprofit Organization	New York	United States	NY	NaN	1172	40.780033	-73.945703	NaN
2	875875415863397	Animal Shelter	New York	United States	NY	10029	98	40.792456	-73.939264	326 E 110th St
3	100423658293123	News & Media Website	NaN	NaN	NaN	NaN	164	NaN	NaN	NaN
4	127949550650477	Local Business	New York	United States	NY	11385	175	40.709530	-73.866167	NaN

STEP 3: Use Python or R to analyze the returned dataframe results. Remember to install libraries before analyzing results.

i **For more information about query syntax, see Meta for Developers page.**

Return to Table of Contents