

gesis

Leibniz-Institut
für Sozialwissenschaften



RESEARCH
FOR THE
DIGITAL AGE



Data Sharing in Social Media Research – Insights to Researchers' Practices and Challenges

Dr. Katrin Weller (GESIS & CAIS)

katrin.weller@gesis.org

General Contexts of Social Media Research

Online platforms such as social media platforms have become crucial elements of our lives – and have thus also become the object of academic research.

The "ABC"

Social media data can also help to identify

Attitudes and opinions,

Behavior,

Characteristics

of human users of digital technologies.

Research based on data
from social media platforms is not a
consistent field.

different platforms
different methods
different disciplines
different motivations
different skills

different opportunities to access
(restricted) data

Researchers' Practices and Experiences

Researchers' Practices and Experiences ...in General

Researchers report similar experiences
when interacting with social media
platforms and the complexities they are
entangled in

(unclear) legal
frameworks

research ethics

changing platform
access options / ToS

rate limits

user expectations /
privacy

Researchers report similar experiences
when interacting with social media
platforms and the complexities they are
entangled in

platforms as
black boxes

data access

interdisciplinarity

methods as
black boxes

data sharing

publishing
practices

ephemerality

missing data

...but arrive at different conclusions when it comes to addressing specific challenges

E.g., in the context of **research ethics**:

- ▶ Big vs. small data
- ▶ Users as authors vs. users as research subjects
- ▶ Particularly vulnerable groups (e.g., activists) vs. professional / public accounts (e.g., politicians)

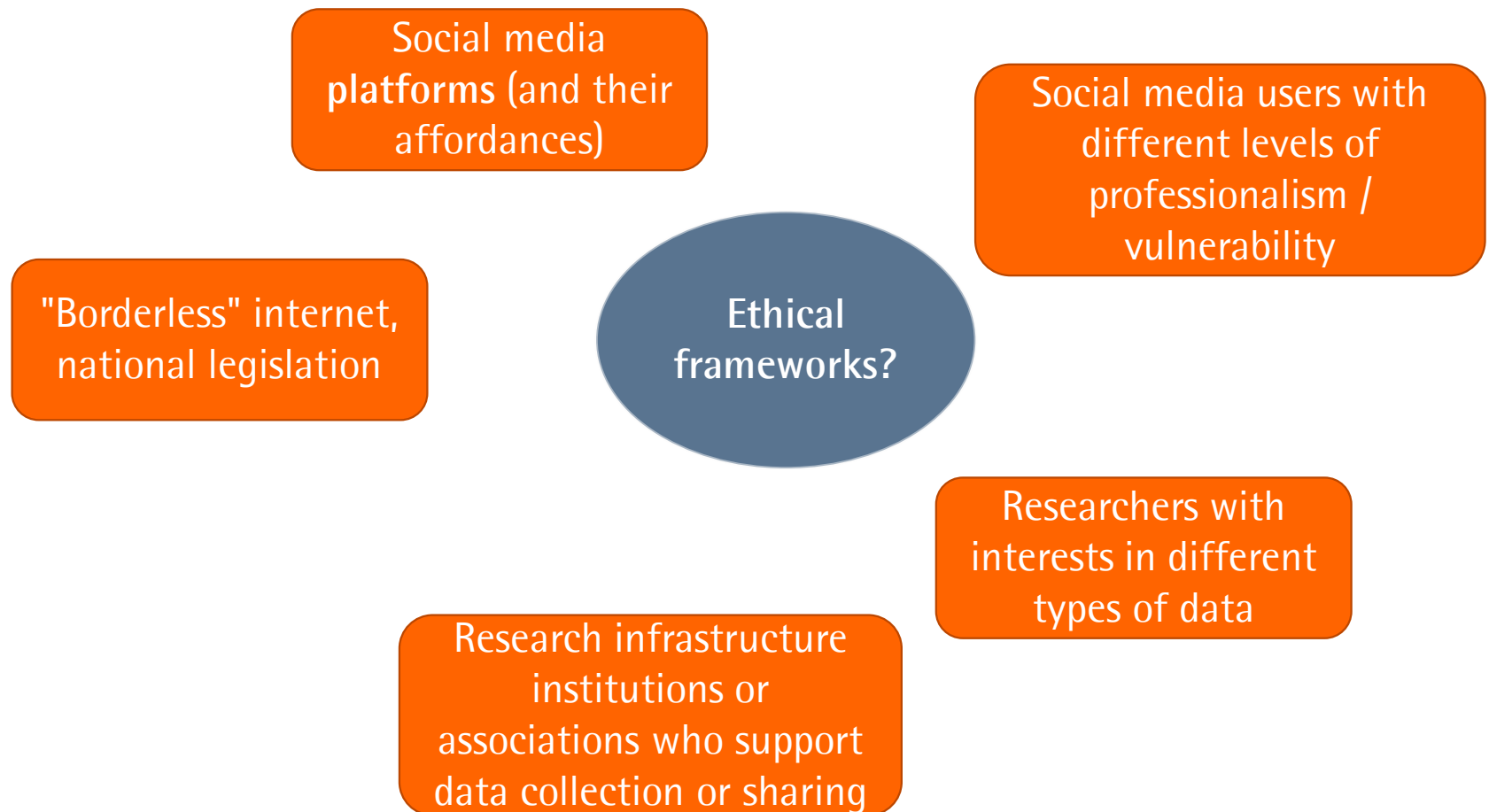
...but arrive at different conclusions when it comes to addressing specific challenges

Or in the context of **data collection and sharing**:

- ▶ Collections via APIs, scraping, copy/pasting, screenshots
- ▶ balancing between following principles of good scientific practice and between respecting legal constraints
- ▶ Not sharing data to protect users vs. sharing to include users

Weller, Katrin, and Katharina E. Kinder-Kurlanda. 2015. "Uncovering the Challenges in Collection, Sharing and Documentation: The Hidden Data of Social Media Research?." In *Standards and Practices in Large-Scale Social Media Research: Papers from the 2015 ICWSM Workshop. Proceedings Ninth International AAAI Conference on Web and Social Media Oxford University, May 26, 2015 – May 29, 2015*, 28-37. Ann Arbor, MI: AAAI Press.

Different entities that effect potential study design – and research ethics



No formal research field - no standard methods

Lack of standards for

- data collection
- methods for preprocessing and analysis
- documentation / data management
- research ethics / data protection

Development of best practices is impacted by the changing nature of social media platforms and users.

Researchers' Practices and Experiences ...in data sharing

Preserve and publish results and data



- Enhance overall research quality by supporting reproducibility and transparency. This includes potentially processed data, e.g. annotations, results from NLP processes.
- Publishing datasets can reduce the need to collect the same kind of data for different research projects.
- Several practical challenges often prevent efficient data sharing. Ongoing challenges for research infrastructure institutions.
- Extra need to care for data protection.

social media researchers are often willing to share data


- Perceived ethical obligations towards the scientific community and towards social media users
- "rich gets richer" phenomenon
- enable better peer review and quality standards



"But you can't make your data available for others to look at, which means both your study can't really be replicated and it can't be tested for review. But also it just means your data can't be made available for other people to say: Ah you have done this with it, I'll see what I can do with it"

Weller, Katrin, and Katharina E. Kinder-Kurlanda. 2015. "Uncovering the Challenges in Collection, Sharing and Documentation: The Hidden Data of Social Media Research?." In *Standards and Practices in Large-Scale Social Media Research: Papers from the 2015 ICWSM Workshop. Proceedings Ninth International AAAI Conference on Web and Social Media Oxford University, May 26, 2015 – May 29, 2015*, 28-37. Ann Arbor, MI: AAAI Press.

... and not everyone is happy to reuse existing datasets

 *"I actually only use [other researchers' datasets] where I'm very sure about where it comes from and how it was processed and analyzed. There is too much uncertainty in it."*

Weller, Katrin, and Katharina E. Kinder-Kurlanda. 2015. "Uncovering the Challenges in Collection, Sharing and Documentation: The Hidden Data of Social Media Research?." In *Standards and Practices in Large-Scale Social Media Research: Papers from the 2015 ICWSM Workshop. Proceedings Ninth International AAAI Conference on Web and Social Media Oxford University, May 26, 2015 – May 29, 2015*, 28-37. Ann Arbor, MI: AAAI Press.

Challenges

sources for challenges

Legal and ethical challenges

- ethical challenges and users' expectations
- the platforms' Terms of Services
- proprietary data
- other legal aspects (e.g. copyright)

Archiving challenges

- lack of standard metadata and procedures
- Setups for searching for existing datasets
- Long-term preservation

(Diversity of) social media platforms and their data

- variety of social media data
- ephemerality: the ever changing nature of social media platforms and their users

Social media research

- (researchers motivations for sharing)

sources for challenges

Legal and ethical challenges

- **ethical challenges and users' expectations**
- **the platforms' Terms of Services**
- **proprietary data**
- **other legal aspects (e.g. copyright)**

Archiving challenges

- lack of standard metadata and procedures
- Setups for searching for existing datasets
- Long-term preservation

(Diversity of) social media platforms and their data

- variety of social media data
- ephemerality: the ever changing nature of social media platforms and their users

Social media research

- (researchers motivations for sharing)

sources for challenges

Legal and ethical challenges

- ethical challenges and users' expectations
- the platforms' Terms of Services
- proprietary data
- other legal aspects (e.g. copyright)

Archiving challenges

- lack of standard metadata and procedures
- Setups for searching for existing datasets
- Long-term preservation

(Diversity of) social media platforms and their data

- variety of social media data
- **ephemerality: the ever changing nature of social media platforms and their users**

Social media research

- (researchers motivations for sharing)

Ephemerality?



1. PLATFORM



2. USERS



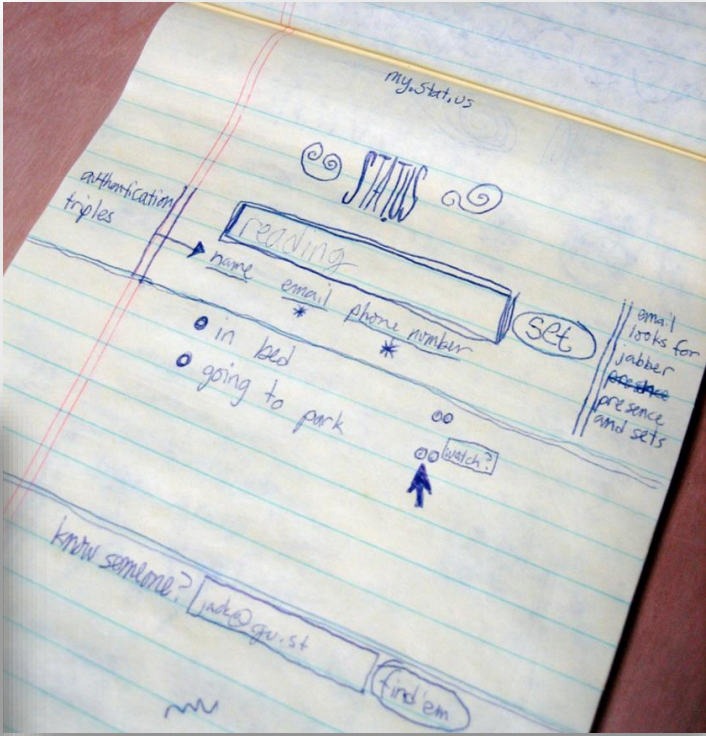
3. CONTENT



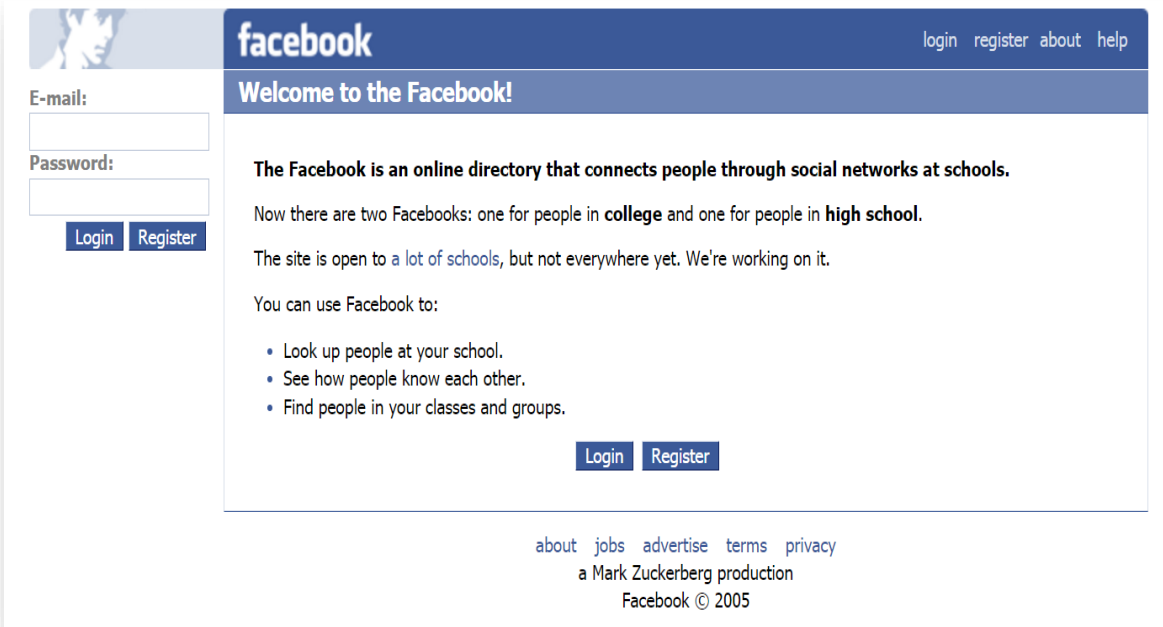
4. DATA/ACCESS

1. Platforms: interfaces, look and feel

Challenge: Long term preservation of social media, i.e. archiving of data and as well as of social media platforms' look and feel.



<https://www.flickr.com/photos/jackdorse/182613360>



Facebook on Dec. 25, 2005. Via Internet Archive's Wayback Machine

2. lost context: user demographics

User statistics

- Per country
- Over time
- For different demographic groups (age, gender, etc.)

2. Changing user behaviour

Users

- May adopt new ways of using a platform
- May move from one platform to another
- change user names / profiles

3. Content

- User-generated content:
 - ▶ Texts, images, videos, mixed formats
- Process-generated:
 - ▶ Connections (friends, followers, links)
 - ▶ Actions (likes, favs, comments, downloads, edits, searches)
- Metadata + data enrichments:
 - ▶ Timestamps, geo-codes, user profiles
 - ▶ Annotations
 - ▶ External links, websites

Lost context?

#jan25

4. Data Access / Platform Lockdowns?


SOCIAL SCIENCE ONE

Building Industry-Academic Partnerships

CONTACT US

Q

HOME
About Us ▾
Our Facebook Partnership ▾
People
Blog
FAQ



Social Science One implements a new type of partnership between academic researchers and private industry to advance the goals of social science in understanding and solving society's greatest challenges. The partnership enables academics to analyze the increasingly rich troves of information amassed by private industry in responsible and socially beneficial ways. It ensures the public maintains privacy while gaining societal value from scholarly research. And it enables firms to enlist the scientific community to help them produce social good, while protecting their competitive positions.

Social Science One's inaugural project on "the effects of social media on democracy and elections" offers researchers privacy-preserving access to Facebook data; [funding from seven nonprofit foundations](#) from across the ideological spectrum (pooled and regranted by the Social Science Research Council, [SSRC](#)); and a new type of [peer pre-review](#) feedback service designed to speed scientific progress and shorten time-to-publication.

The two-part organizational structure that makes Social Science One possible is described in [this paper](#). A commission of senior academics, including some who sign confidentiality agreements and forego grants and publication, acts as a trusted third party with full access at and a partnership with the company. The commission identifies relevant datasets, writes a series of Requests For Proposals, and is obligated to report to the public any violation of

Computational research in the post-API age

Deen Freelon

University of North Carolina at Chapel Hill

Forthcoming in *Political Communication*

Keywords: API, computational, Facebook, Twitter, social media

2018-08-20

On April 4, 2018, the post-API age reached a milestone. On that day, Facebook closed access to its Pages API, which had allowed researchers to extract all posts, comments, and associated metadata from public Facebook pages (Schroepfer, 2018). This decision followed the company's April 2015 closure of its public search API, which provided searchable access to all public posts within a rolling two-week window (Facebook, n.d.). The closure of the Pages A

 / Developers / Forums

Twitter Developers

Sign Up

Log In



The [v1.1 statuses/filter endpoint](#) will be deprecated on **March 9, 2023**. We encourage you to start using the [filtered stream endpoint](#) in the Twitter API v2 at this time. Read the full announcement [here](#)

Announcing the deprecation of v1.1 statuses/filter endpoint

Announcements twitter-api



TwitterDev Twitter Staff

Jan 9

Jan 9

1 / 2
Jan 9

Today, we are announcing the deprecation of the [v1.1 streaming statuses/filter endpoint](#) ³²⁶, with a formal deprecation date of **March 9, 2023**. This completes the full retirement of v1.1 streaming endpoints (we previously [announced the other components in April 2022](#) ¹⁵⁹). We encourage you to begin the migration process to the [filtered stream endpoint](#) ³⁸⁴ in the Twitter API v2 at this time.

Additionally, beginning today, new client applications will no longer be able to gain access to v1.1 statuses/filter. Developers with client apps already using these endpoints will maintain access until the functionality is retired.

Why we're deprecating this endpoint

We know that deprecations can be challenging and we want to be transparent about what's driving this change. Maintaining this v1.1 endpoint requires energy that can best serve our developer community if

4. Platform governance changes

The way platforms govern user activities has tremendous consequences for how our increasingly digital society is organised. This archive enables citizens, researchers and journalists alike to understand how this governance by platforms has changed over time.

PLATFORM GOVERNANCE ARCHIVE 1.0

An archive of key policy documents by Makin

Facebook
YouTube
Twitter
Instagram
Vimeo

Learn more

<https://pga.hiig.de/>

Where and how to share/archive social media data?

Some solutions

preserving social media data

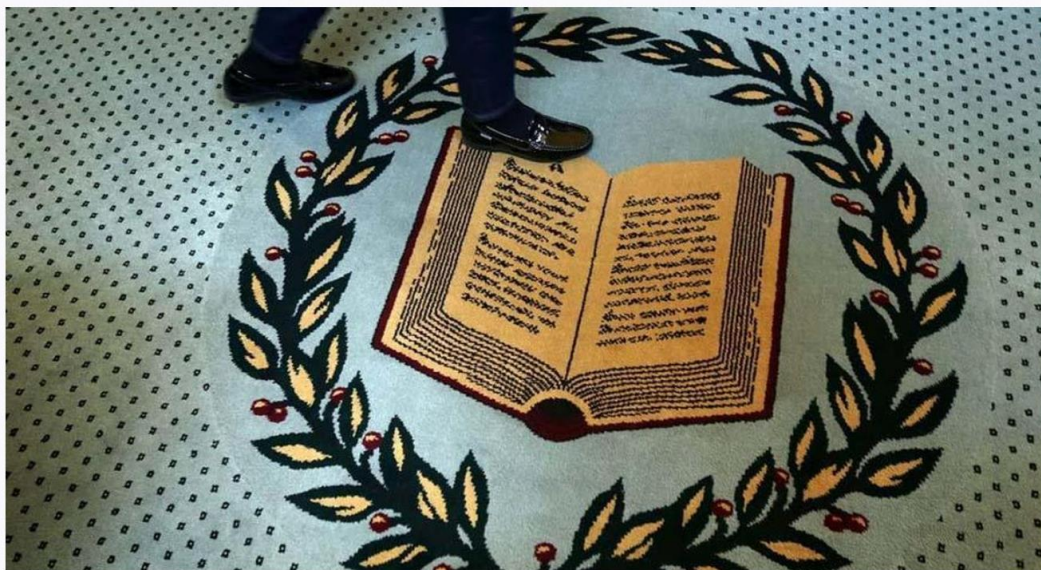


TECHNOLOGY

Library of Congress' Twitter archive is a huge #FAIL

More than five years on, the library's Twitter archive project is in limbo — with no end in sight.

By NANCY SCOLA | 7/11/15 5:09 PM EDT



conferences + journals

Example ICWSM

Paper	Description	# of files	# of observations (tweets/facebook accounts/entries)	# of Twitter users	# of nodes	# of edges
Z Luo, Miles O and T Wang. <i>Opinion Retrieval in Twitter.</i>	Tweets tagged as relevant or irrelevant for 50 specific queries	1	5000			
L Chen, W Wang, M Nagarajan, S Wang and AP Sheth. <i>Extracting Diverse Sentiment Expressions with Target-dependent Polarity from Twitter.</i>	Tweets describing movies and persons	4	426660			
J Mahmud, J Nichols, and C Drews. <i>Where is This Tweet From? Inferring Home Locations of Twitter Users.</i>	Geo tagged tweets from 100 top cities	2	1005259			
L Rossi and M Magnani. <i>Conversation practices and network structure in Twitter.</i>	Tweets about the fifth edition of the popular TV show XFactor Italia	2	22287			
LM Aiello, M Deplano, R Schifanella, and G Ruffo. <i>People are Strange when you're a Stranger: Impact and Influence of Bots on Social Networks.</i>	Social info from anobii.com	7	671585			
J Park, M Cha, H Kim, and J Jeong. <i>Managing Bad News in Social Media: A Case Study on Domino's Pizza Crisis.</i>	Tweet collection for Domino's pizza crisis	10	4645			
Y He, C Lin, W Gao,						

OBTAINING DATASETS

Download and sign the [ICWSM Dataset Usage Agreement](#). Please note that this agreement gives you access to all ICWSM-published datasets. In it, you agree not to redistribute the datasets. Furthermore, ensure that, when using a dataset in your own work, you abide by the citation requests of the authors of the dataset used.

Email the signed agreement, as a PDF file, to dataset-request@icwsml.org. In the body of your email,

1. Be clear that you are requesting access to the ICWSM datasets
2. Include your name,
3. your email address, and
4. the name of your organization.

We will respond to your request with a URL, a username, and a password with which you can download the datasets. **Please allow seven business days for a response.**

individual solutions and grey markets



If people ask us for a data set because they are working on something similar by email, then we might share it, but we can never publicly do that.

The Twitter Project Page at MPI-SWS

Welcome to our Twitter project page. The data used in our ICWSM'2010 paper is available for use by the wider community. Based on Twitter's explicit request, we are only sharing the anonymized topology of the Twitter social network. Please understand that we are not allowed to share any tweet information. **If you have any questions please send us an email at [twitter-contact \(at\) mpi-sws.org](mailto:twitter-contact@mpi-sws.org).**

<http://twitter.mpi-sws.org/>

researchers

- Individual researchers are sharing their datasets e.g. on personal websites, github, figshare...
- Groups of researchers starting collections of datasets

The screenshot shows the CrisisLex.org website. At the top, there is a navigation bar with links: Home, Download Collections, Create a Collection, Download Lexicons, and Create a Lexicon. The main heading is "CrisisLex.org". Below the heading, a paragraph explains that social media is used for crisis information and that CrisisLex.org is a repository of crisis-related social media data and tools. It includes links for "collections of crisis data" and "a lexicon of crisis terms". Below this, there are two buttons: "Collections of Crisis Data »" and "Lexicons of Crisis Terms »". A "References" section follows, listing three academic papers with their authors, titles, and publication details.

researchers / institutes

- Research institutes / research groups starting (topical) data collections



(SOCIAL) MEDIA OBSERVATORY

Mit dem (Social) Media Observatory wird eine Datenbasis zur systematischen Beobachtung medienbasierter Öffentlichkeit aufgebaut, die andere Projekte im Forschungsinstitut Gesellschaftlicher Zusammenhalt für ihre Projekte nutzen können.

Ziel dieses Projektes im Rahmen des **Forschungsinstituts gesellschaftlicher Zusammenhalt (FGZ)** ist der Aufbau eines (Social) Media Observatory (SMO). Als zentrale organisatorische Einheit soll das SMO innerhalb des gesamten FGZ fünf wesentliche Aufgaben erfüllen:

1. die Konzeption und Umsetzung einer technischen Infrastruktur, bestehend aus cloud-basierten virtuellen Servern, einer umfassenden Datenbank gesellschaftlicher Akteure und Organisationen und eigens entwickelten Open Source-Skripten und frei zugänglichen Softwarepaketen, welche im gesamten Projektzeitraum für die a) kontinuierliche und akteursbezogene und b) anlass- und einzelfallbezogene systematische Beobachtung medienbasierter Öffentlichkeit genutzt wird (INFRA);
2. die Erstellung und Pflege einer deutschsprachigen „Handreichung“, die in Form eines Wikis auf bestehende Datensätze, Tools, Softwarepakete o.ä. im Bereich der (social) media observation verweist (DOKU);
3. die Entwicklung eines Fortbildungsangebots mit regelmäßigen halbjährlichen Workshops

Bedarf unterstützt, z. B. durch nach Bedarf terminierte und projektbasierte Sprechstunden (BERA):

5. die Zusammenarbeit mit externen institutionellen Partnern (GESIS, Social Science One), um die sichere, zuverlässige und langfristige Archivierung der innerhalb des Konsortiums gewonnenen (Social) Media-Daten zu gewährleisten und so deren Nachnutzungspotenzial zu erschließen (ARCH).

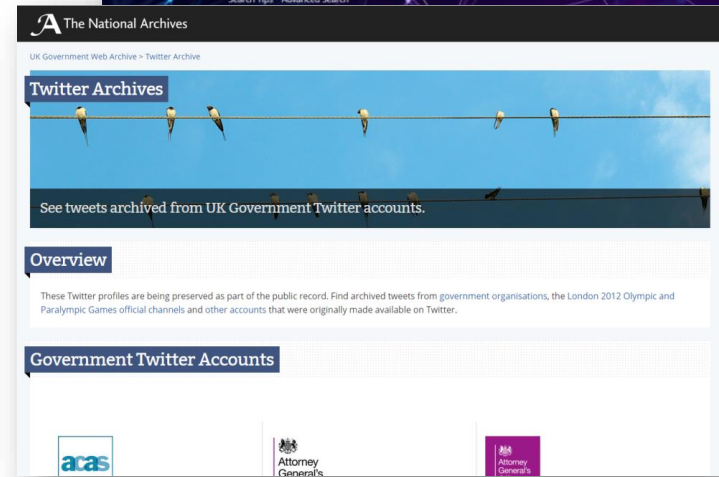
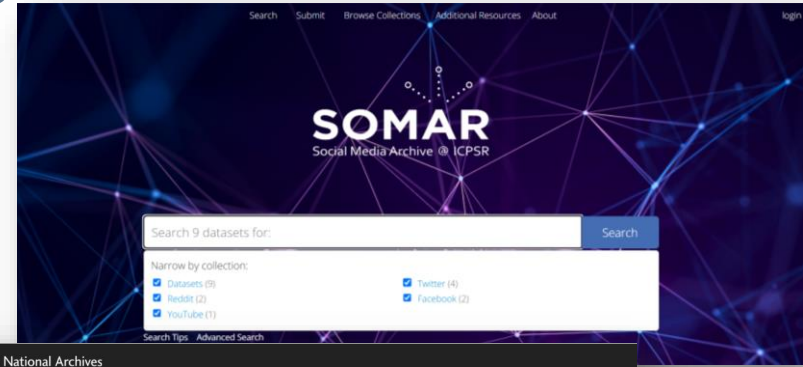
Das Projekt verfolgt somit keine eigene inhaltliche Fragestellung, sondern hat die Aufgabe, allen interessierten Projekten des FGZ Daten, Werkzeuge und Kompetenzen zur Verfügung zu stellen, um eigenständig Fragestellungen mit Bezug zu medienbasierter Öffentlichkeit bearbeiten zu können.

- zur [Projektseite \(Social\) Media Observatory](#)
- zur [Projektbeschreibung auf der Website des FGZ](#)

archives

Institutions with background
in archiving material

- ICPSR (SOMAR)
- UK National Archives
- GESIS
- Deutsche Nationalbibliothek



Hemphill, L., Leonard, S. H., & Hedstrom, M. (2018). Developing a Social Media Archive at ICPSR. In Proceedings of Web Archiving and Digital Libraries (WADL'18). doi:[10.475/123_4](https://doi.org/10.475/123_4)

Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*, 4(2).

Examples from GESIS

Digital Behavioral Data: Datasets

Social scientists are increasingly drawing on web data to analyze social behavior, opinion formation, cultural preferences, or political polarization. Collecting social media data and other digital behavioral data (DBD) up to the standards of social science research is a non-trivial task and often a challenge to individual researchers. GESIS develops innovative methods for the collection of digital behavioral data in the social sciences. In accordance with the proprietary and privacy restrictions that apply, we provide the resulting data for scientific re-use. GESIS offers a range of collected, curated, and augmented datasets; these data are transparent, ready-to-use and often accompanied by additional materials or tools. We concentrate on topical data relevant for the social sciences, training data – e.g., for attribute or opinion detection – or large datasets that can be further mined for individual research purposes.



German Federal Elections

Topical Collection

Source: Twitter, Facebook

These datasets present results from the social media monitoring of Facebook and Twitter for the German federal election campaigns 2013, 2017, and 2021. The project collects the tweets and Facebook posts of political candidates and organizations and the engagement of users with these contents.

[2013 Data](#) | [2017 Data](#) | [2013 Report](#) (637 kB) | [2017 Report](#) | [2021 Data](#) | [2021 Report](#) | [Tool](#) | [Paper](#) | [MTE Talk](#)

TweetsCOVID19

Longitudinal Crawl

Source: Twitter

Semantically annotated corpus of tweets related to the COVID-19 pandemic capturing online discourse about various aspects of the pandemic and its societal impact from October 2019 onwards. The dataset contains precomputed entity and sentiment annotations and extracted tweet metadata. The data are publicly available.

[Description](#) | [Report](#) | [Data](#)

'Call me sexist but' (CMSB)

Topical Collection, Training Data

Source: Twitter, Crowdsourced

The 'Call me sexist but' dataset (CMSB) is part of our work to analyze different dimensions of sexism in social media, including overt hostile sexism, 'benevolent' sexism, or more subtle forms that pose a particular challenge for automatic detection techniques. With this we aim at improving methods for, e.g., addressing sexism on online platforms.

[Data](#) | [Paper](#) | [Blog](#) | [Github](#) | [Github](#)

Just Another Day on Twitter

Platform Data, Baseline Data

Source: Twitter

The dataset "Just another day on Twitter" presents a complete dump of one day on Twitter (September 20-21, 2022), generated by a globally co-ordinated effort from 80 scholars. Being "just another day" the 24 hours covered fall into a turbulent period with Elon Musk about to acquire the platform.

[Data](#) | [Paper](#) | [Blog](#)

VACOS-NLQ

Topical Data, Natural Language Data

Source: Crowdsourced

The natural language query dataset (VACOS-NLQ) is a collection of 3540 written queries for e-commerce product search (laptops and jackets). The queries are enriched with information about age, gender, and domain knowledge of the participants.

[Data](#) | [Source code](#) on GitLab | [Paper](#)

Incels Forum Data

Topical Data, Platform Data

The dataset consists of all publicly visible posts and the data that comes with each post of the online forum incels.is during one week in Nov 2022. The data allows to investigate questions surrounding involuntary celibates, hate-speech, communication in online forums, the emergence of acts of terrorism, and suicide prevention.

[Data](#)

<https://www.gesis.org/en/services/finding-and-accessing-data/digital-behavioral-data-datasets>

Examples from GESIS

31 Hits

Research data
(31)

Variables
(0)

Instruments & Tools
(0)

Publications
(0)

GESIS Library
(0)

GESIS Webpages
(0)

What do you find in this category?

Filter results

Digital Behavioral Data ✕

[Topic](#) ▼
 [Author](#) ▼
 [Publication year](#) ▼
 [Geography](#) ▼
 [Source](#) ▼
 [Study title](#) ▼
 [Study group](#) ▼
 [Collection year](#) ▼
 [Thematic collection](#) ▼

[show more filters](#)

[Careers in Electronic Dance Music \(2001-2015\)](#)

[Jadidj, Mohsen](#)

. Data File Version 1.0.0, <https://doi.org/10.7802/2360>

Abstract: Electronic Dance Music (EDM) is a music style as well as a global creative industry. Performing live and releasing music are the main practices in this field. This datase ... [more](#)

[Inhaltsanalyse von Responsibility Frames zu Gesundheitsthemen in sozialen Medien](#)

[Wiedicke, Annemarie](#)

. Data File Version 1.0.0, <https://doi.org/10.7802/2503>

Abstract: Es werden hier das Codebuch und die Daten einer standardisierten Inhaltsanalyse der Depressions- und Diabetesposts im sozialen Medium Instagram zur Verfügung gestellt (N ... [more](#)

[Invasion@Ukraine](#)

Publ. Jopina Markmann, Simon Assenmacher, Dennis

Sort by:

Relevance ▼

Downloads
[Datasets](#)

Actions
[Bookmark](#)
[Cite](#)

Downloads
[Datasets](#)
[Codebook](#)

Actions
[Bookmark](#)
[Cite](#)

Downloads
[Datasets](#)

Potential solutions

Benchmark-Datasets?

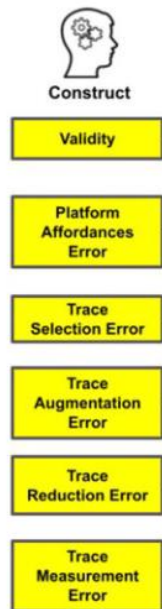
The screenshot shows the Zenodo interface. At the top, there is a search bar and navigation links for 'Upload' and 'Communities'. The main content area displays the dataset title 'Individual Edit Histories of All References in the English Wikipedia' with a date of 'July 31, 2020'. Below the title, the authors are listed: Zagovora, Olga; Ulloa, Roberto; Weller, Katrin; Flöck, Fabian. A description states: 'This dataset includes the historical versions of all individual references per article in the English Wikipedia. Each reference object also contains information about its original creating editor, editors implementing changes to it, and'.

The screenshot shows a dataset page for 'Just Another Day on Twitter: A Complete 24 Hours of Twitter Data' by Pfeffer, Jürgen. It includes a DOI link (https://doi.org/10.7802/2516), an abstract, and a list of authors and their institutions. A sidebar on the right contains links for 'Downloads', 'Aktionen', 'Merken', and 'Zitieren'. The page also lists the study number (SDN-10.7802-2516), DOI (10.7802/2516), publication year (2023), current version (1.0.0), availability (Freier Zugang), and collection period (2022-09-20).

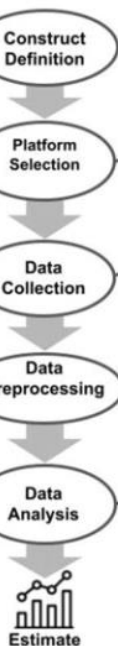
Documentation?

Potential solutions

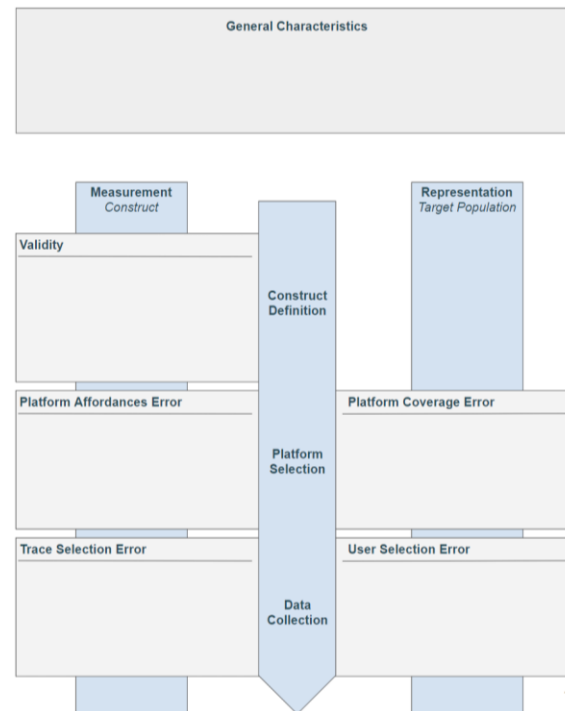
MEASUREMENT



REPRESENTATION



TES-D "Dataset Name"



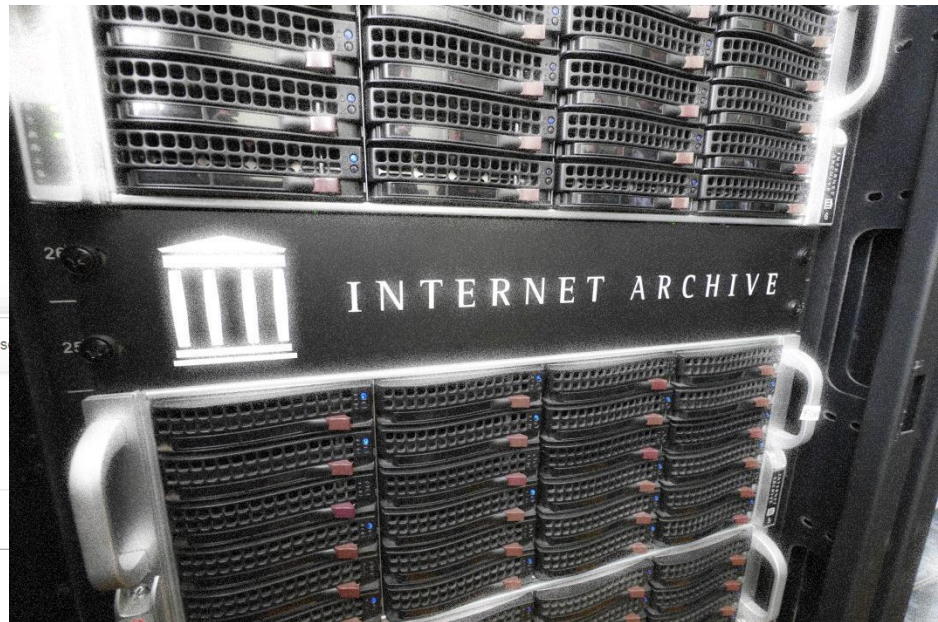
Sen, I., Flöck, F., Weller, K., Weiß, B., & Wagner, C. (2021). A Total Error Framework for Digital Traces of Human Behavior on Online Platforms. *Public Opinion Quarterly* 85 (S1): 399–422. doi: <https://doi.org/10.1093/poq/nfab018>.

Total Error Sheets for Datasets (TES-D).
Work in progress bei GESIS (preprint on arxiv to be published next week)

archives

New Players

- Internet Archive
- Dataverse



HARVARD
Dataverse

Search ▾ About Us

Metrics 4,890,997 Downloads

twitter [Advanced Search](#)

Dataverses (6)

Datasets (82)

Files (134)

Dataverse Category

Research Project (3)

Researcher (2)

Metadata Source

Harvard Dataverse (212)

Harvested (10)

Publication Year

2017 (58)

2014 (40)

2018 (40)

2016 (34)

2015 (17)

[More...](#)

Subject

Social Sciences (67)

Computer and Information Science (15)

Arts and Humanities (5)

Earth and Environmental Sciences (3)

Business and Management (3)

1 to 10 of 222 Results

Trends In Twitter Hashtag Applications.pdf

Apr 1, 2014 - Trends in Twitter hashtag applications
Adobe PDF - 274.6 KB - MD5: 97cb7141e6476d0b0fae4a6a1cb531bc

Twitter

Trends in **Twitter** hashtags.
Filename Without Extension: Trends in **Twitter** Hashtag Applications
File Tag: **Twitter**

Twitter-Replication.R

Jun 4, 2018 - Replication Data for: Launching Revolution: Social Media and the Egyptian Uprising's First Movers
R Syntax - 8.6 KB - MD5: e217d39f20f543f666d8d29160111dd7

Filename Without Extension: **Twitter-Replication**

twitter-code-A.ipynb

Apr 18, 2017 - Replication Data for: "Effects of Investor Sentiment Using Social Media on Corporate Financial Distress" dissertation
Unknown - 12.3 KB - MD5: 5a5da7fb0a56dbbee5e8e7ecc5af05fa0

Filename Without Extension: **twitter-code-A**

django-twitter-import-checkpoint.ipynb

Apr 18, 2017 - Replication Data for: "Effects of Investor Sentiment Using Social Media on Corporate Financial Distress" dissertation
Unknown - 2.2 MB - MD5: af1b00cc2ab734a41da8ae2333ed1066

Filename Without Extension: django-**twitter**-import-checkpoint

twitter-code-A-checkpoint.ipynb

Apr 18, 2017 - Replication Data for: "Effects of Investor Sentiment Using Social Media on Corporate Financial Distress" dissertation

overview: examples of practices

Approach	Example	Retrievability	Documentation + standards	Long-term perspective
„Grey market“	Data shared with colleagues (often upon request)	Low , via personal connections	Low , no standardized documentation or data format	Low , no guaranteed long term availability
Researchers' personal/ professional websites		Medium , URLs may be referenced	Low , no standardized documentation or data format	Low , no guaranteed long term availability
Social media providers	e.g. Wikipedia dumps	High , if directly provided from social media platform	Medium/high , depending on the producer	Medium/high , depending on the producer
Project-based or thematic collections	e.g. KONECT, CrisisLex	Medium , URLs may be referenced	Medium , if same principles are applied within entire collection	Low , no guaranteed long term availability
Conferences & journals	e.g. ICWSM datasets with conference papers	High , usually related to accepted publications	Medium/high , depending on the publisher	Medium/high , depending on the publisher
Professional archives	e.g. datasets at GESIS data archive	Medium/high , datasets may be referenceable with DOIs.	Medium/high , depending on the archive's requirements	High , guaranteed availability for different time spans

- For additional examples see our paper: <http://dx.doi.org/10.1145/2908131.2908172>
- See also: Thomson, S.D. 2016. Preserving Social Media. DPC Technology Watch Report. Retrieved from <http://dpconline.org/publications/technology-watch-reports>

New hope?



Log in English

Search on Europa

Search

Law

Delegated Regulation on data access provided for in the Digital Services Act

Have your say > Published initiatives > Delegated Regulation on data access provided for in the Digital Services Act > Feedback and statistics: Call for evidence

Unique feedback (133)

Statistics

Order by

Most recent



Showing results 1 to 10

31 May 2023 | Academic/research Institution

Psychology of Technology Institute (United States)

Thank you for the opportunity to provide input on how to best provide data access in the public interest for very large online platforms. We would like to provide one very specific suggestion:

Mandate access to existing relevant product experimentation data. This is important for 3 reasons: -

https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/feedback_en?p_id=32045757



News | May 31, 2022

EDMO releases report on researcher access to platform data

The report by the European Digital Media Observatory's (EDMO) Working Group on Platform-to-Researcher Data Access includes a draft Code of Conduct on how platforms can share data with independent researchers while protecting users' rights.



Institute for Data,
Democracy & Politics
THE GEORGE WASHINGTON UNIVERSITY

**Report of the European Digital Media Observatory's Working Group
on Platform-to-Researcher Data Access**

31 May 2022

<https://edmo.eu/2022/05/31/edmo-releases-report-on-researcher-access-to-platform-data/>

Next Steps

Conclusions and next steps

- Changing platform landscape and changing legal situations require lots of resources for adjusting to new circumstances.
- Critical reflections, ethical considerations and research data management need to be built into the entire research process.
- (More) joint efforts to enable data quality standards.
- Technical solutions for (remote) secure access, anonymization, aggregation are needed.

This can only be achieved together

Networking?

- NFDI
- RatSWD
- DGPuK
- Web Archiving
Community (z.B.
netpreserve)
- CESSDA
- Who else?

Thank you and virtual greetings from Düsseldorf/Cologne/Bochum

Questions are welcome:

Dr. Katrin Weller

Team lead „Digital Society Observatory“
Computational Social Science
GESIS Leibniz Institute for the Social Sciences
katrin.weller@gesis.org

Co-Team lead „Research Data and Methods“
Center for Advanced Internet Studies (CAIS)
katrin.weller@cais-research.de