

# Social Media Data for Research. Ethical Implications

June, 8 2022

Frederic Guerrero-Solé  
Universitat Pompeu Fabra

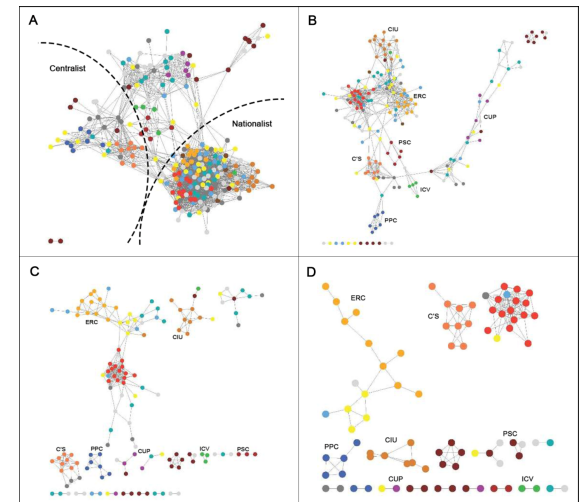
## Frederic Guerrero-Solé

Polarization, duplication of active audiences and populism on Twitter. Analysis of the influence of populist discourses on political debates in Spain (2016-2020). funded by the Spanish Ministry of Science and Innovation. PGC2018-097352-A-I00.



## Latest researches

- ❑ Guerrero-Solé, Frederic, López-González, Hibai (2019). *Government Formation and Political Discussions in Twitter. An Extended Model for Quantifying Political Distances in Multiparty Democracies*. Social Science Computer Review.
- ❑ Guerrero-Solé, F. (2018). Interactive behavior in political discussions on Twitter. Politicians, media and citizens' patterns of interaction in the 2015 and 2016 electoral campaigns in Spain. Social Media and Society.



## Latest researches

- ❑ Guerrero-Solé, Frederic; Philippe, Olivier (2020). «The toxicity of Spanish politics in Twitter during the COVID-19 pandemics». *Hipertext.net*, 21, 133-139.
  - ❑ Guerrero-Solé, Frederic (2021). *The Ideology of the Spanish Media. A Measure of the Political Leaning of Spanish News Media based on Twitter Users' Interactions and People's Perception of Parties' Ideology.*
-

## Methodologies

- ❑ Anonymization and privacy. Focus on the individual.
  - ❑ Collecting data: storage and context. Informed consent.
  - ❑ Accountability, public data and inequalities.
  - ❑ Analysis and reporting/publishing results. Latent attributes.
  - ❑ Interventions.
  - ❑ Experience from DATAPOP and other questions.
-

Questions on the ethical implications in research with social media data

**Just Because it is Accessible Doesn't Make it Ethical!**

Starting from Six Provocations for Big Data (boyd & Crawford, 2011)

## Anonymization

Just Because it is Accessible Doesn't Make it Ethical (boyd & Crawford, 2011)

**De-anonymization** and compromising of the privacy of people not aware their data has been collected (Zimmer 2008).

## Anonymization

- ❑ Is it possible to anonymize data collected from social media? How?

## Research questions

- ❑ A relevant reflection: does it matter to know who did/published what? Only for certain cases (public personae). **Research questions according to this.**
  - ❑ My advice: **never report results that allow other to de-anonymize** (do not put the focus on **users' identities**, unless it is the main goal of the research, and related to the previous point).
-



## Informed consent in social networks. Is it possible?

- Practices like 'informed consent' and protecting the privacy of informants are intended to empower participants in light of earlier abuses in the medical and social sciences (Blass, 2004; Reverby, 2009).
- It may be unreasonable to ask researchers to obtain consent from every person who posts a tweet, but it is unethical for researchers to justify their actions as ethical simply because the data is accessible.
- Researchers must keep asking themselves – and their colleagues – about the ethics of their data collection, analysis, and publication.

## Storage and context

- Do we have (are we allowed) to storage data from social media?
  - The legal limitations. And the limitations imposed by the platforms.
  - How can data be interpreted in the future? It's not possible to manage the removal of removed content.
  - Storage: under certain circumstances.
  - What is the access researchers have to data? Protected? Encrypted?
-

## Storage and context

- Deleted accounts
- How should deleted accounts/posts be dealt with? (i.e., deleted account = withdrawn consent?)

## Context collapse

- What if someone's 'public' blog post is taken out of context and analyzed in a way that the author never imagined?
- What does it mean for someone to be spotlighted or to be analyzed without knowing it?
- Who is responsible for making certain that individuals and communities are not hurt by the research process?

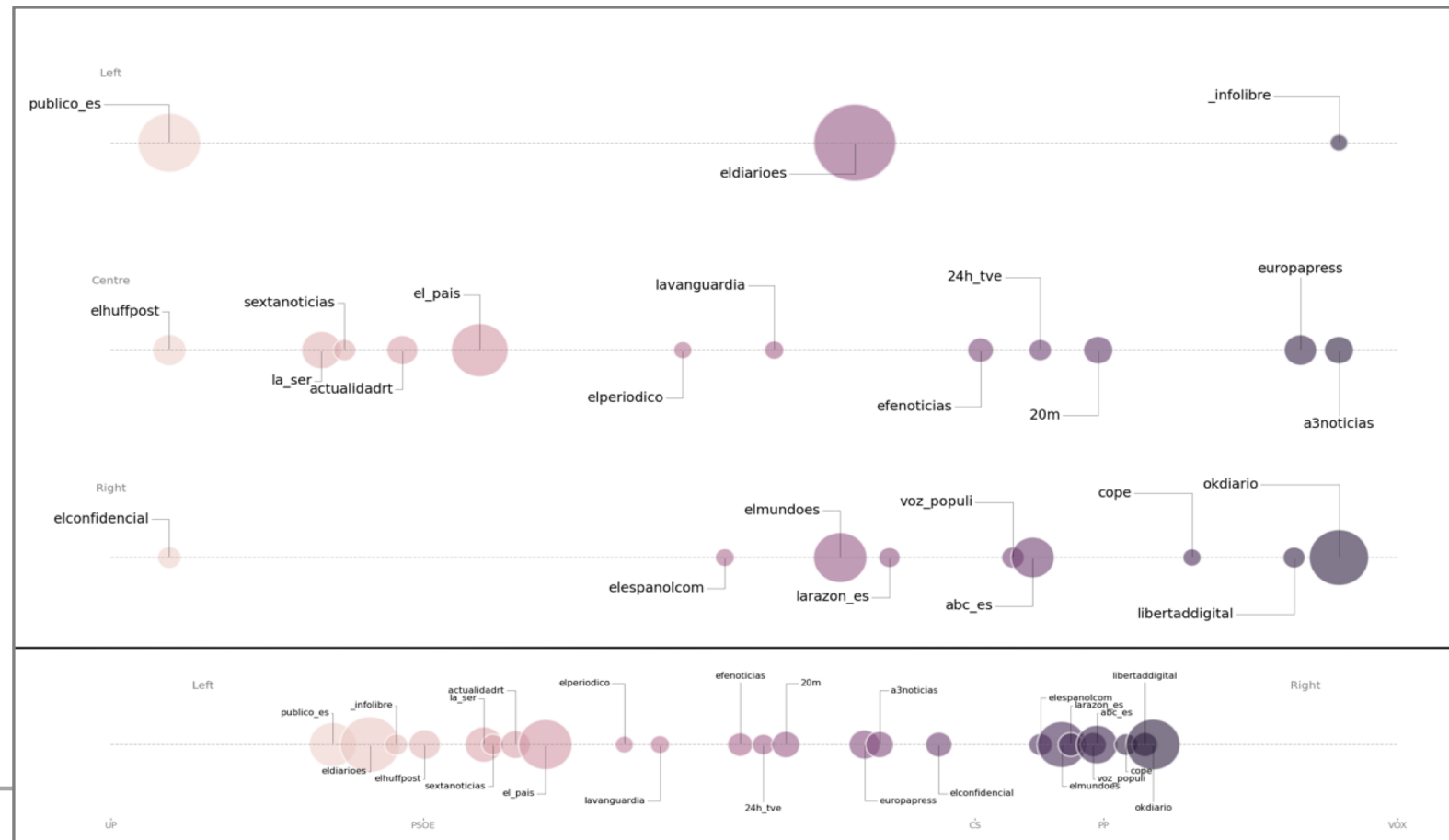
## Accountability

- Importance of accountability: accountability to the field of research, and accountability to the research subjects.
- Many ethics boards do not understand the processes of mining and anonymizing Big Data, let alone the errors that can cause data to become personally identifiable.

## Public data (?) and inequalities of access

- There are significant questions of truth, control and power in Big Data studies: researchers have the tools and the access, while social media users as a whole do not.
- Data may be public (or semi-public) but this does not simplistically equate with full permission being given for all uses.
- There is a considerable difference between being in public and being public, which is rarely acknowledged by Big Data researchers.

# Analysis and reporting/publishing results.



## Analysis and reporting/publishing results

- Classification and sensitive categories.
- Uncovering users' latent attributes? Such as ideology, sexual orientation, religion?
- You can report the handles of politicians and political parties, media outlets, but not journalists. Discussion.
- What if reviewers ask you to report personal data, such as handles, usernames, and other?



## Intervention

- Even when researchers try to be cautious about their procedures, they are not always aware of the harm they might be causing in their research. For example, a group of researchers noticed that there was a correlation between self-injury ('cutting') and suicide.
- They prepared an educational intervention seeking to discourage people from engaging in acts of self-injury, only to learn that their intervention prompted an increase in suicide attempts. **For some, self-injury was a safety valve that kept the desire to attempt suicide at bay. They immediately ceased their intervention** (Emmens & Phippen 2010).

## Intervention

- Although analysis can lead to the detection of certain problematic issues (suicide, for example), intervention is not ethically justified.
  - The same for political polarisation.
-

## Experience from DATAPOP Project (2019-2022)

- Twitter data gathered by means of Twitter APIs.
- A considerable number of fields collected by user.
- Databases are encrypted, Access only by researchers, password.
- Users are anonymized by means of a simple script.

## Experience from DATAPOP Project (2019-2022)

- Focus ONLY on indirect relationships between users (not even direct, who retweets (follows/like) to whom!).
- No personal private data used in any research.
- Bulk analysis of content.
- Personal Information only when users are politicians and media and political organizations.

## Experience from DATAPOP Project (2019-2022)

- Ethics is technology? Yes, but...
- Above anything, ethics is about scientific attitude and behavior in relation to other people's data.
- Research questions in play? Do they follow the ethical requirements of your organization? (Obviously, legal requirements are out of discussion). And the ethics of the platform?

## Experience from DATAPOP Project (2019-2022)

- Do ethical issues put researchers in an inferior situation in relation to other organizations and companies?
- Are these issues preventing reach interesting scientific results and conclusions?
- How to deal with these differences, inequalities?

## Experience from DATAPOP Project (2019-2022)

- Universities need to adopt ethical standards, but also fight against the unethical use of data by other organizations in society.
- How can we compare the use of these data with the use of human or animal DNA, for example?

## Other questions

- Can social media be used to recruit participants?
- Interacting with social media users for research purposes/using social media to observe vs. condition people's behavior?



## Other questions

- Does the type of social media platform (e.g. Twitter vs. LinkedIn vs. Tinder) make a difference as to how the data should be handled? (Taking into account users' reasonable expectations of privacy)
- The answer is NO. What does **reasonable expectations** mean? Some places are more public than others?

# Social Media Data for Research. Ethical Implications

June, 8 2022

Frederic Guerrero-Solé  
Universitat Pompeu Fabra