

<https://helda.helsinki.fi>

---

## Capturing Emerging Realities in Citizen Engagement in Science in Social Media : A Social Media Analytics Protocol for the Allinteract Study

Pulido Rodriguez, Cristina M.

2021-10-01

---

Pulido Rodriguez , C M , Ovseiko , P , Palomar , M F , Kumpulainen , K & Ramis , M 2021 , ' Capturing Emerging Realities in Citizen Engagement in Science in Social Media : A Social Media Analytics Protocol for the Allinteract Study ' , International Journal of Qualitative Methods , vol. 20 , 16094069211050163 . <https://doi.org/10.1177/16094069211050163>

---

<http://hdl.handle.net/10138/336893>

<https://doi.org/10.1177/16094069211050163>

---

cc\_by\_nc

publishedVersion

---


*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

# Capturing Emerging Realities in Citizen Engagement in Science in Social Media: A Social Media Analytics Protocol for the Allinteract Study

International Journal of Qualitative Methods  
Volume 20: 1–9  
© The Author(s) 2021  
DOI: 10.1177/16094069211050163  
[journals.sagepub.com/home/ijq](https://journals.sagepub.com/home/ijq)  


Cristina M. Pulido Rodriguez<sup>1</sup> , Pavel Ovseiko<sup>3</sup> , Marta Font Palomar<sup>2</sup>,  
Kristiina Kumpulainen<sup>4</sup>, and Mimar Ramis<sup>2</sup>

## Abstract

In the digital era, social media has become a space for the socialization and interaction of citizens, who are using social networks to express themselves and to discuss scientific advances with citizens from all over the world. Researchers are aware of this reality and are increasingly using social media as a source of data to explore citizens' voices. In this context, the methods followed by researchers are mainly based on the content analysis using manual, automated or combined tools. The aim of this article is to share a protocol for Social Media Analytics that includes a Communicative Content Analysis (CCA). This protocol has been designed for the Horizon 2020 project Allinteract, and it includes the social impact in social media methodology. The novel contribution of this protocol is the detailed elaboration of methods and procedures to capture emerging realities in citizen engagement in science in social media using a Communicative Content Analysis (CCA) based on the contributions of Communicative Methodology (CM).

## Keywords

virtual environments, mixed methods, observational research, dimensional analysis, methods in qualitative inquiry

## Background

Citizens are increasingly using social networks for socialization, interaction with other users and expression of ideas and thoughts. Therefore, social media has become a space where citizens' voices are expressed. Statistics show that since 2017, the number of users in social media has increased almost by one billion, reaching a total of 4.2 billion active users in 2021 (Johnson, 2021).

In order to collect citizens' voices, researchers are increasingly using social media to gather information about citizens' perceptions, needs, thoughts or debates. This reality is reflected in the growing number of scientific articles related to social media content analysis. According to the search conducted in Web of Science on June 18th, 2021 using the keywords 'social media' AND 'content analysis', the number of such articles has increased from 10 in 2009 to 710 in 2020.

Research on social media shows that social media can provide a useful source for research with social impact as

researchers can collect citizens' voices to conduct research that improves citizens' lives and respond to society's challenges. For example, Silver and Matthews (2017) showed how citizens used Facebook groups to seek information and support and to self-organize the community in response to the natural disaster caused by a tornado. Similar conclusions on the usefulness of social media for research with social impact

<sup>1</sup>Department of Journalism and Communication Studies, Autonomous University of Barcelona, Cerdanyola del Valles, Spain

<sup>2</sup>Radcliffe Department of Medicine, University of Oxford, Oxford, UK

<sup>3</sup>Department of Sociology, Universitat de Barcelona, Barcelona, Spain

<sup>4</sup>Department of Education Learning, Culture and Interventions, University of Helsinki, Helsinki, Finland

## Corresponding Author:

Cristina M. Pulido Rodriguez, Department of Journalism and Communication Studies, Autonomous University of Barcelona, C/ de les Vinya, Cerdanyola del Valles 08193, Spain.  
Email: [cristina.pulido@uab.cat](mailto:cristina.pulido@uab.cat)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

were reached by [Gálvez-Rodríguez and colleagues \(2019\)](#) [Gálvez-Rodríguez et al. \(2019\)](#) when they analysed citizens' support in social media after terrorist attacks. Research related to health issues has used social media as a source for raising awareness about breast cancer ([Miller et al., 2019](#)), cancer's early diagnosis ([Prochaska et al., 2017](#)) and the COVID-19 pandemic, among others. Related to this last example, several studies have delved into citizens' mental health during the pandemic ([Arasli et al., 2020](#); [Newman et al., 2021](#)), citizens' hesitancy to vaccines ([Hernandez et al., 2021](#); [Piedrahita-Valdés et al., 2021](#)), the responses provided by world leaders ([Rufai & Bunce, 2020](#)) and the impact of fake news ([Atehortua & Patino, 2021](#); [Islam et al., 2020](#); [Pulido, Ruiz-Eugenio, et al., 2020](#)), among other topics.

All these studies provide insightful examples of how social media can be a useful source for research with social impact, as it enables the collection of citizens' voices. However, the studies focused on social media analyses use diverse methods to collect and analyse their data. These differences are mainly related to the social network chosen for the analysis, the procedure for the identification of keywords, the definition of the units of analysis or the process of data analysis. Recently, some authors have developed protocols for content analysis in a specific social network ([Reuter & Lee, 2021](#); [Stens et al., 2020](#)). These protocols provide a step-by-step description of how to conduct content analysis in social media, which is the cornerstone of research replicability. They are based on 1) a top-down approach, in which researchers choose the keywords to be used, and 2) a standard manual coding process in which two independent researchers code the messages extracted, and Cohen's kappa coefficient is calculated to assess reliability.

Yet, researchers that use social media as a source for their research use different methods to conduct the coding and analysis, including manual, automated or hybrid methods. In this debate about the use of manual or automated methods, [Lewis and colleagues \(2013\)](#) pointed out how the implementation of a combined approach of manual and computational methods ensures the rigour and human sensitivity while processing big amounts of data. Although the implementation of hybrid methods brings together the benefits of both methods, they base their interrater reliability on Cohen's kappa. As [Pulido, Villarejo-Carballido and colleagues \(2020\)](#) pointed out, this is a traditional and correct way to assess the reliability of content analysis, but it provides no space for dialogic reliability, in which time, plurality of voices and egalitarian dialogue are essential.

To fill this gap, Social Media Analytics (SMA) and the Communicative Content Analysis (CCA) provide novel approaches to content analysis in social media. Drawing upon the postulates of Communicative Methodology (CM), which acknowledge that 'all individuals have inherent capacities for communication and social interaction and that they can understand the world, generate knowledge and change social structures' ([Gómez et al., 2019](#), p. 3), the cornerstone of this

approach is its dialogic basis. The CM is built upon the premise that dialogic co-creation of knowledge occurs when researchers and citizens engage in egalitarian discussions, where researchers provide scientific evidence, citizens provide their daily life experiences and the dialogue between them takes place with the aim of reaching a consensus through the use of validity claims and not of power claims ([Gómez et al., 2019](#); [Gómez González, 2021](#); [Pulido et al., 2018](#)). Thus, in SMA, all the steps and procedures are discussed and agreed upon using egalitarian dialogue to reach a real consensus that enables further co-creation of knowledge. This communicative app Hernandez roach to social media analysis allows us to move beyond the existing debates about the use of manual, automated or combined methods to provide new knowledge about the application of CM to social media. This CM has widely demonstrated its scientific and social validity in several studies, including studies in the field of social media analysis ([Cabré-Olivé et al., 2017](#); [Oliver et al., 2021](#); [Pulido et al., 2018](#); [Pulido, Ruiz-Eugenio, et al., 2020](#); [Pulido, Villarejo-Carballido, et al., 2020](#)).

The current article makes a novel contribution to content analysis in social media by presenting a protocol for a Social Media Analytics study, which has been elaborated for Allinteract, a research project funded under the EU Horizon 2020 programme ([Flecha & Pulido, 2021](#)). The procedure and steps detailed in this protocol can be potentially used in other research and fields.

## Explanation and Justification of Method

The current XXXX SMA protocol draws on the SMA methodology developed by [Flecha & Pulido 2017](#). As [Cabré and colleagues \(2017\)](#) explain, there are two approaches for conducting SMA: top-down and bottom-up ([Cabré-Olivé et al., 2017](#)). The top-down approach is based on defining a priori a series of keywords related to the objectives of the research project, in order to explore in social media whether and how citizens use these. The bottom-up approach is based on identifying the topics that emerge from the citizens themselves, through analysing the most used keywords and hashtags in social media and online sources, related to the research project's goals. Once these are identified, they are contrasted with the topics and goals of the research project to analyse whether citizens' interests and opinions expressed on social media are covered by the project. The combination of these two strategies allows researchers to have both an overview of what topics receive most attention from citizens in online interactions, as well as a picture of emerging topics that might have not yet been covered by the research goals. Moreover, the combination of these two strategies can be informative for future research ([Cabré-Olivé et al., 2017](#)), as well as for authorities and other institutions. In this way, in line with the communicative framework, instead of relying on experts' opinions to define issues of relevance and social impact, the SMA allows researchers to gather the voices of

different citizens and social agents in social media on the issues which are most relevant to them. As [Cabr  and colleagues \(2017, p. 99\)](#) put it: ‘This strategy is useful for being more connected to the needs and concerns of citizens and can help to refine the research goals according to this information collected’.

Among the current social concerns, the COVID-19 pandemic has been a major topic of debate on social media ([Pulido, Villarejo-Carballido, et al., 2020](#)), and this *infodemic* related to it has been studied by diverse research projects. Previous studies pointed to false information being more likely to be shared ([Vosoughi et al., 2018](#)). However, a study by [Pulido, Villarejo-Carballido and colleagues \(2020\)](#) has shed new light on the matter through the use of the SMA methodology, finding differing trends in the sharing of COVID-19–related false and evidence-based information on Twitter. Indeed, after analysing 942 tweets, the authors found that, whereas false information was more likely to be tweeted, tweets containing evidence-based or fact-checking information were more likely to be retweeted. This methodology provides researchers with a deeper analysis of the reality being studied, contributing essential knowledge for current and future research. Moreover, it can also help authorities better understand citizens’ interactions in social media and act according to citizens’ interests and concerns. In the case of the COVID-19 *infodemic*, for instance, [Pulido, Villarejo-Carballido and colleagues’ \(2020\)](#) findings might be taken up by health authorities to post more tweets from official accounts.

Besides creating deeper knowledge with scientific and political impact, SMA contributes knowledge about how potential or real social impact of research is being shared on social media, both quantitatively and qualitatively ([Pulido et al., 2018](#)). This makes social media platforms relevant for the dissemination of scientific evidence of the social impact of research. Through these platforms, messages transcend the walls of academia and are made accessible for citizens. Indeed, making such evidence available to all through sharing, commenting and engaging with the content is key. This distinguishes Social Impact in Social Media (SISM), a subset of SMA, from other methodologies aimed at measuring social impact, as it differentiates dissemination of research results from social impact of research ([Pulido et al., 2018](#)). Thus, SISM puts forward a new methodology that allows capturing which scientific advancements citizens share in social media and thus find relevant. Among others, the use of this methodology has allowed researchers to capture the fact that the number of tweets and Facebook posts containing evidence of social impact does not depend on the number of tweets or posts made by certain research projects, but on how many of these tweets or Facebook posts contain scientific evidence ([Pulido et al., 2018](#)). As an example, the most recent application of this methodology has been in the field of health in a study aimed at analysing the types of interactions in which citizens contribute to spreading misinformation and which

types of interactions promote the overcoming of fake news around health on social media ([Pulido, Ruiz-Eugenio, et al., 2020](#)). Results revealed that interactions containing misinformation are mostly aggressive, whereas interactions based on evidence of social impact are transformative and respectful. Moreover, through the use of SISM on Reddit, Twitter and Facebook, [Pulido and colleagues \(2018\)](#)[Pulido et al. \(2018\)](#) were able to identify that messages based on evidence of social impact overcome misinformation. By including citizens’ voices in this bottom-up approach, SISM provides public health professionals with strategies to overcome fake news around health issues, as well as with relevant knowledge to design interventions to promote citizens’ discussion of scientific evidence of social impact on social media (2020).

Following the CM and its aim of contributing to achieving social impact, egalitarian dialogue is a key aspect of the analysis in SMA. Beyond quantitative data, which is usually obtained from analyses of social media, the CCA within SMA requires a profound human analysis of social media interactions. To that end, researchers engage in a continuing dialogue, from the development of the codebook for analysis to the end of the research project (2020). This dialogue allows the unveiling of new findings that emerge from the interaction itself and that could not have been reached by the simple addition of the individual analyses of the researchers involved. Moreover, such an approach contributes to the robustness of the findings, as it is presented in the *Rigour section* below.

## Sampling/Recruitment

### Selection of the Social Networks

The main aim of the Allinteract SMA is to include the diversity and plurality of citizens’ voices and perspectives. Thus, it was essential to conduct the research in those social networks that citizens use, considering the diverse options and spaces where users can interact in each social network. For this reason, considering the exploration of the different social networks included in the Allinteract SMA protocol ([Flecha & Pulido, 2021](#)), the selected sources to include citizens’ voices in SMA in the top-down strategy were Facebook pages, Twitter and Instagram hashtags and Reddit communities.

### Procedure and Criteria of selection

This protocol includes a twofold strategy, including a top-down and a bottom-up approach. Each of the approaches has followed a different procedure, each of which is detailed below.

**Top-Down Strategy.** As defined by [Cabr -Oliv  et al. \(2017\)](#), in the top-down approach to SMA, the keywords, pages and hashtags are defined by researchers according to the aims of the research. In the case of the current SMA protocol, the procedure was discussed dialogically by a diverse group of

researchers in order to collect a plurality of voices and perspectives. Table 1 shows the procedure carried out under each social network, as well as the criteria followed in the final selection.

**Bottom-Up Strategy.** The bottom-up approach is driven by the voices of citizens and aims to directly identify and collect the topics that emerge from the keywords that are most used in citizens' messages on social media. This strategy was implemented on Twitter, and daily Trending Topics in all European countries were used to identify those that were relevant for the project. In order to identify Trending Topics, several tools were used:

1. Websites that collect Trending Topics in each hour in the different European Countries,
2. Twitter Trending Topics in the different European Countries,
3. *Python* Software.

Researchers collected the lists of the Trending Topics and explored the content of the messages included within each hashtag in order to identify those that were related to the aims of the research. Table 2 compiles the number of hashtags explored, the number of hashtags related to the scope of the project and the final number of hashtags included in the sample.

In order to select the final hashtags to be included in the sample, the following criteria were considered:

1. Relation with the aims of the research: in this case, the hashtags to be selected needed to be related to gender or education.
2. Language: considering the international team of researchers that constitutes the project and participated in the analysis, the tweets needed to be mainly in English language.
3. Representation of different European countries: since the SMA aimed to include the voices of European citizens from diverse backgrounds, the presence of hashtags emerging from different European countries was important. In the same line, hashtags that were

**Table 1.** Procedure and Criteria for the Selection of Top Pages/Hashtags.

	Procedure of search	Number of keywords, hashtags or pages explored	Criteria	Indicator
Twitter	<ol style="list-style-type: none"> <li>1. Search most used hashtags on the topic</li> <li>2. Search each proposed hashtag</li> <li>3. Include 'top' and 'latest' tweets</li> </ol>	Hashtags: 37	Voices of citizens  Popularity of the hashtag High level of interaction among users	High presence of profiles from individual citizens rather than companies and organizations High use of the hashtag recently High number of retweets and replies
Facebook	<ol style="list-style-type: none"> <li>1. Combine two keywords in the browser</li> <li>2. Refine by 'pages' and include all locations and categories</li> </ol>	Combinations of keywords: 22 Number of pages explored: 54	High activity of the page High level of interaction with users Outreach of the page Type of content shared	High frequency of posts shared High number of comments in each post Number of followers Promotion of debates, presence of scientific evidence
Instagram	<ol style="list-style-type: none"> <li>1. Search keywords in the browser (general keywords may be more used and contained in other hashtags)</li> <li>2. Select most popular hashtags containing the keyword</li> <li>3. Search all posts using the selected hashtags</li> </ol>	Combinations of keywords: 13 Number of hashtags explored: 91	Popularity of the hashtag Voices of citizens  High level of interactions among users	Recent high use of the hashtag High presence of profiles from individual citizens rather than companies and organizations Number of comments and likes under each post
Reddit	<ol style="list-style-type: none"> <li>1. Search keyword in the browser</li> <li>2. Refine by 'Communities and Users'</li> </ol>	Keywords: 8 Communities: 29	Type of content shared in the community Size of the community High level of interactions among users	Posts that promote debate among users rather than humorous and anecdotal Number of members Number of comments and interactions under each post

**Table 2.** Summary of the Identification of Hashtags in the Bottom-Up Approach.

Date	Number of trending topics analysed	Number of trending topics related to the scope of the research	Number of hashtags included in the sample
08/03/21	591	57 hashtags	0
09/03/21	1242	21 hashtags	3
10/03/21	1233	17 hashtags	3
11/03/21	1273	9 hashtags	2
12/03/21	1306	6 hashtags	1
13/03/21	1246	4 hashtags	2
14/03/21	1311	4 hashtags	0
15/03/21	1313	6 hashtags	0
TOTAL	9515	124	11

Trending Topic at European level or in more than one country were given priority.

- Representation of diverse topics: Citizens use more than one hashtag in each tweet. Therefore, several of the hashtags identified in the Trending Topics were included in the same messages, which were related to recent news or cases. In order to avoid duplicate messages, one of the criteria was to include hashtags related to diverse topics.

### Selection of the Sample

Following the top-down and bottom-up strategy, researchers extracted the messages posted during the study period of time using NVIVO and PYTHON. The sample of this analysis included a maximum of 10,000 messages under each of the 40 hashtags or pages in the top-down approach and of the 11 hashtags emerged from the bottom-up approach. [Table 3](#) compiles the criteria followed to select the final messages that constituted the sample.

Altogether, the final sample included a total of 79,352 messages in the gender topic and 111,881 messages in the education topic.

### Data Handling/Analysis

The process of handling and analysing the collected data consisted of four phases: the definition of the units of analysis, the design of the dialogic codebook, dialogic coding and the CCA.

#### Definition of Units of Analysis

The first step was to establish the criteria for the coding and analysis of the collected information. These criteria were agreed by all researchers and decided by consensus:

- The unit of analysis included the complete tweet or post with links, images or videos attached.

**Table 3.** Criteria for the Selection of the Sample.

Criteria of Inclusion	Criteria of Exclusion
To have been posted in the established period of time	Not meeting the selection criteria
To have been posted by a real user (not a bot)	Not meeting the scope of the project
To include the selected hashtags or keywords or to be posted in the selected pages	
To be among the 10,000 messages with more number of interactions among users	

- The number of retweets and likes were part of the analysis.
- Only messages posted by real public users were taken into account. Spam, fake accounts, bots and private users were discarded.
- In cases of multiple levels of hyperlinks, at least two levels (i.e. a link contained inside another link in the posted message), were accessed, checked and considered.

### Design of the Dialogic codebook

First, researchers designed the dialogic codebook for the CCA basing all the decisions on egalitarian dialogue. Regular dialogic discussions were scheduled to address each of the points in the codebook. Therefore, all decisions in the design of the codebook were agreed upon and made using validity claims instead of power claims. The final codebook used for the coding of all the datasets can be found in the protocol elaborated for the project ([Flecha & Pulido, 2021](#)). In addition, researchers conducted a pilot to test the process and the codebook. The pilot was conducted in Facebook and Twitter and included two Facebook pages and two Twitter hashtags obtained with a top-down strategy, as well as one Twitter hashtag obtained from a bottom-up approach. The extracted data was coded by a diverse group of researchers, who followed the draft version of the dialogic codebook. All the decisions were taken on the basis of an egalitarian dialogue. The conclusions of the pilot were shared in the regular dialogic

discussions, in order to identify possible improvements in the codebook and to ensure its rigour and appropriateness. The pilot provided insightful messages that exemplified each of the codes and provided further clarification for the coding process.

### *Dialogic Coding of the Data*

In the dialogic coding of data, researchers worked with the entire dataset and coded each message according to the dialogic codebook. The dialogic coding was carried out by a diverse team of researchers that included researchers from six institutions from 6 European countries. The inclusion of researchers from diverse profiles, areas of knowledge and countries provided a plurality of voices that enriched the discussions and the co-creation of knowledge.

The coding process included two steps: (1) the assignment of a code 0–5 according to the aims of the research and (2) the categorization of the sources used in the messages, depending on the presence or absence of scientific evidence. The coding process was developed manually:

1. First, all messages were distributed among the researchers in the coding team, and one researcher carefully read the units of analysis and categorized the messages following the codebook. With the information contained in the message, the researcher assigned a number 0–5 according to the aim of the research and a number 0–2 depending on the sources referred.
2. Second, a double check process was carried out by a different researcher, who re-read and re-coded all the messages. When researchers had doubts about particular messages or differences between the previous and the new code, the message was dialogically discussed and all researchers in the coding team provided their arguments in order to reach an agreement based on validity claims.

During the coding process, a new category in the use of scientific sources emerged. Researchers realized that citizens used social media interactions to request the scientific evidence supporting other users' arguments and also used scientific arguments to discuss the validity of the sources provided. Therefore, researchers reviewed the emerging findings and decided to include a new code 3 in the coding of scientific evidence.

### *Communicative Content Analysis*

Once all messages were coded, all data was analysed quantitatively and qualitatively. Researchers in charge of the analysis carefully read all the units of analysis under each objective and filled the analysis grids according to the information provided. Just like in the case of the design of the codebook, dialogue has been at the core of the design of the analysis grids.

*Qualitative Analysis.* For the qualitative analysis, a Communicative Content Analysis (CCA) was conducted. CCA was first conducted by [Pulido, Villarejo-Carballido and colleagues \(2020\)](#), following the postulates of the Communicative Methodology ([Gómez et al., 2019](#)). Following this communicative approach, the analysis included two dimensions:

1. Transformative dimension: including messages that provided evidence of the factors that facilitate social, political and scientific impact of research, or of the elements that promote the implementation of the targeted actions and policies.
2. Exclusionary dimension: including messages that showed the obstacles that hinder the achievement of the social, political and scientific impact of research, or the implementation of the targeted actions and policies.

Each grid was designed specifically to respond to each objective of the research and included categories, such as the type of content (campaign, discussion, awareness-raising action, citizens' mobilization, implementation of actions, etc.), the profile of the promoter of the message (citizens, NGOs, policymakers, enterprises, etc.), the level of intervention (local, national, international, etc.), the topics covered and the evidence of social impact (potential and real), among others. The complete analysis grids can be found in the Allinteract – Social Media Analytics Protocol ([Flecha & Pulido, 2021](#)).

Following CM's postulates which are at the core of CCA, when researchers could not clearly classify a message, this message was dialogically discussed, and researchers reached an agreement based on validity claims. The analysis also included the description of the similarities and differences found between each social network and between the top-down and bottom-up strategy, in terms of topics discussed, scientific evidence shared or interactions among users, among others.

*Quantitative Analysis.* The quantitative analysis was conducted when all the grids were filled in with the information obtained from the units of analysis. The quantitative analysis included a descriptive analysis of the percentage of messages under each category and dimension and also of the presence of scientific evidence in the messages. Other relevant analysis in this regard included the analysis of the kind of messages that got more interaction from citizens, the relationship between the presence of scientific evidence and the number of interactions among users, the presence of real or potential evidence of scientific, political and social impact of research or differences in the presence of real or potential evidence of impact across categories, among others.

### **Ethics**

The design and implementation of the Allinteract project follows fundamental ethics standards to ensure the quality and

excellence during and after the project, including the scientific and ethical procedures defined by the EU Charter of Fundamental Rights and the UNESCO Universal Declaration of Human Rights. Therefore, ethical principles have been at the core of all Allinteract activities, including in the design and elaboration of the Social Media Analytics protocol. The project was approved by the Bioethical Committee of the University of Barcelona, as the institution coordinating the project. Besides, other partner institutions also provided an ethical approval by their institutions, in the cases in which such a measure was foreseen. First, the Allinteract consortium elaborated a Data Management Plan that sets the requirements to be followed in all project activities involving data. In particular, it states that all data collected from social media needs to be anonymized to respect the social media users' privacy rights. In addition, Allinteract researchers have ensured the compliance with the Terms and Conditions of each social network, as well as with the EU GDPR regulation. To ensure compliance with the EU GDPR and with the Terms and Conditions of each social media company, all extracted data was anonymized and messages posted by fake accounts or bots were discarded.

## Rigour

The approach to qualitative rigour adopted in the current Allinteract SMA protocol is based on the application of four quality criteria for the qualitative research: credibility, transferability, dependability and confirmability (Korstjens & Moser, 2018). The first criterion – credibility (Lincoln & Guba, 1985) – refers to the confidence of the research findings. In the case of the SMA protocol, credibility based on the interpretation of raw data extracted from the social media networks is respectful of the meaning of messages' users. There are different strategies for ensuring this criterion, for instance, prolonged engagement and persistent observation, which were afforded by a pilot study. This pilot study invested sufficient time to become familiar with the data and context for testing the analysis' categories and to improve the coding process.

The data triangulation was conducted using multiple data sources, in this case, different social media (Twitter, Facebook, Reddit, Instagram) via two different strategies (bottom-up and top-down), as explained in previous sections. In addition, investigator triangulation was conducted through a dialogic process that involved researchers from six institutions from six European countries involved in the process of the CCA. The latter followed the methods and procedures developed by Pulido et al. (2018). The dialogic process between researchers consisted of different actions: (a) the dialogic codebook was created and agreed by all the researchers, (b) a permanent online forum in the virtual team workspace was set up to allow discussion of questions and concerns regarding the coding of particular messages, (c) when researchers had doubts about the coding of a particular message, this message was discussed with the coordinator team. In such cases, the relevant messages were highlighted with different colours and

researchers provided arguments for assigning one code or another. This way, this dialogue between the researchers and coordinator team ensured that the final code is based on arguments and (d) to ensure consistency in coding, researchers from the coordinator team reviewed all the messages analysed by the other research teams of all institutions involved. The version of the coding reviewed and agreed by the coordinator team was shared in the virtual team workspace with the other research teams, and they were invited to discuss it further. The final version of the coding was agreed by the coordinator team and the other research teams by consensus.

The second criterion is transferability, based on the strategy of thick description (Roller & Lavrakas, 2015). The Allinteract SMA protocol and data analysis explains this strategy in detail, the sample size used, the different strategies for data extraction, the type of data deleted to comply with the ethical criteria, and the type of the data kept for the purpose of analysis. In line with this strategy, the results of the analysis are explained within a context that could be understood by any kind of reader in order to facilitate the transferability judgement as required in this criterion.

The third and fourth criterion – dependability and confirmability – can be unified, as indicated by Korstjens and Moser (2018), through the strategy of audit trail (Tracy, 2010). In line with this strategy, all the research steps are described transparently to ensure the confirmability step by step from the research design to the reporting of the findings. The Allinteract SMA protocol itself helps to ensure dependability and confirmability, where the details are described in-depth with illustrative examples. Moreover, all the research decisions are made available on the project workspace through sharing documents with agreements reached and forum posts that keep the record of each step taken.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was developed within the research project ALLINTERACT. Widening and Diversifying Citizen Engagement in Science that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement num 872396.

## ORCID iDs

Cristina M. Pulido Rodriguez  <https://orcid.org/0000-0001-8630-7529>  
Pavel Ovseiko  <https://orcid.org/0000-0002-3504-2177>

## References

Arasli, H, Furunes, T, Jafari, K, Saydam, MB, & Degirmencioglu, Z (2020). Hearing the voices of wingless angels: A critical content



- analysis of nurses' COVID-19 experiences. *International Journal of Environmental Research and Public Health*, 17(22), 8484. <https://doi.org/10.3390/ijerph17228484>.
- Atehortua, N. A., & Patino, S. (2021). COVID-19, a tale of two pandemics: Novel coronavirus and fake news messaging. *Health Promotion International*, 36(2), 524–534. <https://doi.org/10.1093/heapro/daaa140>.
- Cabré-Olivé, J., Flecha-García, R., Ionescu, V., Pulido, C., & Sordé-Martí, T. (2017). Identifying the relevance of research goals through collecting citizens' voices on social media. *International and Multidisciplinary Journal of Social Sciences*, 6(1), 70–102. <https://doi.org/10.17583/rimcis.2017.2652>.
- Flecha, R., & Pulido, C. (2017). Definition of Social Media Analytics. This work is licensed under a Creative Commons Attribution- Non Commercial-No Derivatives 4.0 International License. Available in: <https://archive.org/details/socialmedianalytics>
- Flecha, R., & Pulido, C. (2021). Allinteract-Social Media Analytics Protocol is licensed under a Creative Commons Attribution-Non Commercial-ShareAlike 4.0 International License. Available in <https://archive.org/details/sma-protocol-allinteract>
- Gálvez-Rodríguez, M. d. M., Haro-de-Rosario, A., García-Tabuyo, M., & Caba-Pérez, C. (2019). Building online citizen engagement for enhancing emergency management in local European government. *Online Information Review*, 43(2), 219–238. <https://doi.org/10.1108/oir-09-2016-0286>.
- Gómez González, A. (2021). Science with and for society through qualitative inquiry. *Qualitative Inquiry*, 27(1), 10–16. <https://doi.org/10.1177/1077800419863006>.
- Gómez, A., Padrós, M., Ríos, O., Mara, L.-C., & Pukepuke, T. (2019). Reaching social impact through communicative methodology. Researching with rather than on vulnerable populations: The Roma case. *Frontiers in Education*, 4(9), 1–8. <https://doi.org/10.3389/educ.2019.00009>.
- Hernandez, R. G., Hagen, L., Walker, K., O'Leary, H., & Lengacher, C. (2021). The COVID-19 vaccine social media infodemic: Healthcare providers' missed dose in addressing misinformation and vaccine hesitancy. *Human Vaccines & Immunotherapeutics*, 17(9), 2962–2964. <https://doi.org/10.1080/21645515.2021.1912551>.
- Islam, M. S., Sarkar, T., Khan, S. H., Mostofa Kamal, A.-H., Hasan, S. M. M., Kabir, A., Yeasmin, D., Islam, M. A., Amin Chowdhury, K. I., Anwar, K. S., Chughtai, A. A., & Seale, H. (2020). COVID-19-related infodemic and its impact on public health: A global social media analysis. *The American Journal of Tropical Medicine and Hygiene*, 103(4), 1621–1629. <https://doi.org/10.4269/ajtmh.20-0812>.
- Johnson, J. (2021). *Global digital population as of January 2021*. Statista. <https://www.statista.com/statistics/617136/digital-population-worldwide/>.
- Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, 24(1), 120–124. <https://doi.org/10.1080/13814788.2017.1375092>.
- Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, 57(1), 34–52. <https://doi.org/10.1080/08838151.2012.761702>.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE. <https://play.google.com/store/books/details?id=2oA9aWlNeoC>.
- Miller, CA, Guidry, JPD, & Fuemmeler, BF (2019). Breast cancer voices on pinterest: Raising awareness or just an inspirational image? *Health Education & Behaviour: The Official Publication of the Society for Public Health Education*, 46(2\_suppl), 49–58. <https://doi.org/10.1177/1090198119863774>.
- Newman, KL, Jeve, Y, & Majumder, P (2021). Experiences and emotional strain of NHS frontline workers during the peak of the COVID-19 pandemic. *The International Journal of Social Psychiatry*. <https://doi.org/10.1177/00207640211006153>.
- Oliver, E., Llasat, M. C., Llasat-Botija, M., & Díez-Palomar, J. (2021). Twitter's messages about hydrometeorological events. A study on the social impact of climate change. *Sustainability*, 13(6), 3579. <https://doi.org/10.3390/su13063579>.
- Piedrahita-Valdés, H, Piedrahita-Castillo, D, Bermejo-Higuera, J, Guillem-Saiz, P, Bermejo-Higuera, JR, Guillem-Saiz, J, Sicilia-Montalvo, JA, & Machío-Regidor, F (2021). Vaccine hesitancy on social media: Sentiment analysis from June 2011 to April 2019. *Vaccines*, 9(1), 28. <https://doi.org/10.3390/vaccines9010028>.
- Prochaska, J. J., Coughlin, S. S., & Lyons, E. J. (2017). Social media and mobile technology for cancer prevention and treatment-american society of clinical oncology. *American Society of Clinical Oncology Educational Book*, 37, 128-137. [https://doi.org/10.1200/EDBK\\_173841](https://doi.org/10.1200/EDBK_173841).
- Pulido, C. M., Redondo-Sama, G., Sordé-Martí, T., & Flecha, R. (2018). Social impact in social media: A new method to evaluate the social impact of research. *PloS One*, 13(8), e0203117. <https://doi.org/10.1371/journal.pone.0203117>.
- Pulido, CM, Ruiz-Eugenio, L, Redondo-Sama, G, & Villarejo-Carballido, B (2020). A new application of social impact in social media for overcoming fake news in health. *International Journal of Environmental Research and Public Health*, 17(7). <https://doi.org/10.3390/ijerph17072430>.
- Pulido, C. M., Villarejo-Carballido, B., Redondo-Sama, G., & Gómez, A. (2020). COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. *International Sociology*, 35(4), 377–392. <https://doi.org/10.1177/0268580920914755>.
- Reuter, K., & Lee, D. (2021). Perspectives toward seeking treatment among patients with psoriasis: protocol for a Twitter content analysis. *JMIR Research Protocols*, 10(2), e13731. <https://doi.org/10.2196/13731>.
- Roller, M. R., & Lavrakas, P. J. (2015). *Applied qualitative research design: A total quality framework approach*. Guilford Publications. <https://play.google.com/store/books/details?id=0zAXBgAAQBAJ>.
- Rufai, S. R., & Bunce, C. (2020). World leaders' usage of Twitter in response to the COVID-19 pandemic: A content analysis. *Journal of Public Health*, 42(3), 510–516. <https://doi.org/10.1093/pubmed/fdaa049>.
- Silver, A., & Matthews, L. (2017). The use of Facebook for information seeking, decision support, and self-organization following a

- significant disaster. *Information, Communication & Society*, 20(11), 1680–1697. <https://doi.org/10.1080/1369118X.2016.1253762>.
- Stens, O., Weisman, M. H., Simard, J., & Reuter, K. (2020). Insights from Twitter conversations on lupus and reproductive health: Protocol for a content analysis. *JMIR Research Protocols*, 9(8), e15623. <https://doi.org/10.2196/15623>.
- Tracy, S. J. (2010). Qualitative quality: Eight “Big-Tent” criteria for excellent qualitative research. *Qualitative Inquiry*, 16(10), 837–851. <https://doi.org/10.1177/1077800410383121>.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>.