EUMEPLAT
European Media Platforms: assessing positive and negative externalities for European culture

Deliverable D2.1

A Framework and Methodological Protocol for analyzing the platformization of news

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004488

Document information

Grant Agreement #:	**101004488**
Project Title:	**EUROPEAN MEDIA PLATFORMS: ASSESSING POSITIVE AND NEGATIVE EXTERNALITIES FOR EUROPEAN CULTURE**
Project Acronym:	**EUMEPLAT**
Project Start Date:	**01/03/2021**
Related work package:	WP2 Fake news: platformization of Journalism
Related task(s):	T2.1: A methodological framework for analyzing platform journalism
Lead Organisation:	P9 – ISCTE-IUL
Author(s):	Gustavo Cardoso Cláudia Álvares José Moreno Rita Sepúlveda Miguel Crespo Caterina Foà
Status	Final
Submission date:	17/12/2021
Dissemination Level:	Public

Table of Contents

1 Introduction

The goal of this document is to explicit both the procedures implemented in order to design a framework for studying the platformization of news in Europe (WP2, task 2.1) and the tasks that are due for the research of the issue in the 10 countries included in the EUMEPLAT project (WP2, task 2.2). The document will develop in two stages. First describing in detail the tests made for the development of the framework and the choices taken in its design. And, second, describing in detail the envisaged implementation of the framework, in order to respond to the research questions. The first part has to do with design, the second with implementation.

The goal of Work Package 2 is to study “fake news” and the platformization of journalism (between months 1 and 20). Task 2.1 and 2.2 of this WP refer specifically to the “Platformization of News in Ten Countries” (between months 1 and 16). This document is the deliverable 2.1 and is due in month 6 of the EUMEPLAT PROJECT. The research regarding the “Platformization of News in Ten Countries” should start now and its deliverable is due in month 16 of this project, which provides 10 months for its implementation.

In designing this framework, we followed “Platformization” and “Europeanization” as the main guidelines, aiming for the study of how information about Europe and Europeans' main concerns is published and debated on the main social media platforms in the 10 countries participating in the EUMEPLAT project. The research questions are: 1) Which are the most relevant issues in European media, and how are citizens debating about them?; and 2) Which debate is taking shape at the intersection of top-down [professional] and bottom-up [non-professional] communication in social media platforms, in the ten countries? To address these research questions we devised a quantitative method for extracting a significant sample of social media posts and publications on the dimensions to study and a qualitative method for its analysis. The four main dimensions in analysis are **Europe** and the three issues of most concern for European citizens according to Eurobarometer (2020), as they relate to Europe: **Health**, **Economy** and **Environment**.

In order to extract the posts that will compose the sample for analysis (procedure described in detail later in this document), we established a search query similar in the 10 countries (for each of the four themes) and collected a list of relevant news media outlets in each country. To extract data, we chose the main social media platforms in Europe: Facebook, Twitter and YouTube (Statcounter, 2021). By crossing the four search queries with each country and each social media platform, filtered by professional or non-professional content - we obtain six outputs for each country and each dimension: a) Professional news content on Facebook pages; b) User-Generated content on Facebook pages; c) User-Generated content on Facebook public groups; d) Professional news content on Twitter; e) User-Generated content on Twitter; f) User-Generated content on YouTube. According to the framework we will code up to 10 posts or publications on each of the four dimensions on each of the six extractions, which means up to 720 posts or publications for each extraction in each country.

The framework focuses on a period of three months, extracting posts and publications for one month.

This research does not aim to be diachronic but synchronic. Its goal is not to portray the evolution of the debate about the given issues during a given timeframe but rather to establish the current state of the debate about the four dimensions in Europe in the three social media platforms analysed.

Regarding the criteria for choosing the posts to be analysed, this framework chooses to follow a criteria of relevance within social media, therefore adopting the metrics that the social media platforms themselves provide that better suit that goal (again, details for this choice will be developed further on this document). In this sense, posts to extract will be ordered by interactions (on Facebook), by estimated reach (on Twitter) and by relevance (on YouTube). By adopting these metrics (which are the best available in each social media platform as a proxy for relevance) we expect that the framework will be able to identify posts that were influential on the debate of the given dimensions each month, in each country. The posts extracted, of course, are those that correspond to the aforementioned query in each language (details on the construction of the queries further on this document).

With these six extractions for each country/language in each of the four dimensions, the aim is to allow cross-analysis between all of them, permitting, for the same timeframe, comparisons between different countries/languages (ten countries, 12 languages), different social media platforms (Facebook pages and groups; Twitter and YouTube) and different dimensions (Europe, Health, Economy, Environment), published by professional news producers or non-professional actors.

2 The platformization of news in Europe

The goal of this framework is to address both “Europeanization” and “Platformization of News”. Of course, both concepts are highly contested and complex to operationalize. The purpose of this framework is precisely to develop an approach to that, proposing a method to research how the issues that are most important for Europeans are distributed and debated on social media platforms, both by news organizations and general users; and what values, concepts or perspectives of Europe are expressed in the way those issues are distributed and debated.

Europeanization is a concept that may have different interpretations and may be subject to diverse approach perspectives. It may be understood as a process of transference of *legitimacy* from national sources of power to the European Union. It may also be interpreted

as a step in the process of globalization, insinuating some form of *post-national organization*. As an alternative, Europeanization may also come to be viewed as the realization of a *pan-european project* that would be the realization of the ultimate national values of the countries that comprise it. At last, Europeanization may also refer to a *modernization process* by which some countries get rid of old habits or social/economic structures to align with Europe. Common to all of these perspectives is the fact that Europeanization always refers to a *process*. One of the goals of this framework is to try to gauge the expression of these European values and processes on the social media platforms upon which news currently predominantly circulate.

Additionally, Europeanization is also a discursive as well as a material concept. Therefore, its manifestations in the debate about Europe on social media platforms - either authored by the news media or the users - will always result from an assemblage of the discursive and the material (Carpentier, 2021). The aim of the framework is also to try to capture that assemblage.

Platformization, on the other hand, can be defined as the penetration of infrastructures, economic processes and governmental frameworks of digital platforms in different economic sectors and spheres of life, as well as the reorganisation of cultural practices and imaginations around these platforms (Poell, Nieborg, van Dijck, 2019). This process is of course also very complex and involves co-related processes of datafication, commodification and algorithmic distribution (Van Dijck, Poell & De Waal, 2018) that highly influence what information Europeans receive and how they interact with it. In particular, recent studies have pointed to social media platforms as the primary gateway for consumers to interact with news in most developed countries (Newman et al., 2021; Pew Research Report, 2021). This means that not only News Media have to reach audiences through those platforms, but also that other actors beyond the mainstream news media also have access to those platforms and are actively producing and distributing information on it. The goal of this framework is to research precisely that.

3 Constructing a methodological framework for analyzing the platformization of news

The methodology used to study and understand the platformization of news, as defined and viewed in the project's work package 2, falls within digital methods. As mentioned by Rogers (2017, p.75) “Broadly speaking, digital methods may be considered the deployment of

online tools and data for the purposes of social and medium research" and in a first approach this is how we see them.

However, the use of digital methods goes even further as they take into account the medium from which data are collected and its specificities. Omena, following Rogers' work, defines digital methods as "a quali-quanti research practice that re-imagines nature, mechanisms and native data to web platforms and search engines to study society" (Omena, 2019, p.6).

Such positioning and consequent definition makes it necessary, at first, to understand the specifics, the digital grammars (e.g. reactions, comments, hashtags, urls) and the affordances of the medium where the research will fall into a medium-specificity logic (Omena, 2019); and, in a second moment, to understand how the approach should be put into practice, how the questioning and consequent collection should be designed in order to understand and contextualize the meaning of the outputs generated by the collection.

Through this paradigm, digital methods are seen as "a distinctive strategy for internet-related research where the Web is considered an object of study for more than online or digital culture only." (Rogers, 2017, p.91) allowing to study dynamics generated from and on platforms of and by users.

It is also an approach to which dynamism and consequent fluidity are associated with both the medium, due to the recurrent updates and changes in the platforms, i.e. the media where the research takes place, and in the users through their participation. Thus, digital methods are associated with a culture of continuous adaptation (Omena, 2019), a way of investigating that is different from traditional methods.

Taking into account the assumption of what digital methods are and the specifics of the approach, it is necessary to reflect on how the investigation can be carried out. Questions such as What kind of study are we going to perform? What tools are we going to use? How about query design? What are the limitations of the method? Those are some of the questions we propose to answer.

What kind of study are we going to perform?

Social media platforms like Facebook, Twitter or YouTube, by the number of users and usage frequency, play a central role in daily lives and in daily activities. According to the most recent data ("Digital 2021", 2021), Facebook has more that 2,7 billion active monthly users worldwide and YouTube over 2,3 billions. Twitter has more than 350 million users worldwide. But, even more relevant, according to the research by the Reuters Institute for the Study of Journalism, a significant percentage of users of social media affirm receiving most of their news through those platforms, 44% on Facebook, 29% on YouTube and 13% on Twitter (Newman et al., 2021). Not only by mediating as also by shaping the information and

consumption diets, social media platforms are a medium for communication, entertainment, or information and a virtual place where the active roles of users have impact in the construction of narratives concerning different subjects.

To understand how society expresses itself in social media we will monit platforms with the aim of identifying narratives about the dimensions being researched and the circulation/distribution of those narratives. The role of users in social media platforms imply that there are different kind of actors that contribute for the construction and dissemination of narratives while reorganizing communication flows (Coromina, O. & Molina, A.P., 2018).

Monitoring platforms for analysing the platformization of news in the context of europeanization will allow us to identify most relevant posts and operationalize a content analysis following four steps of the methodological framework (see Figure 1).

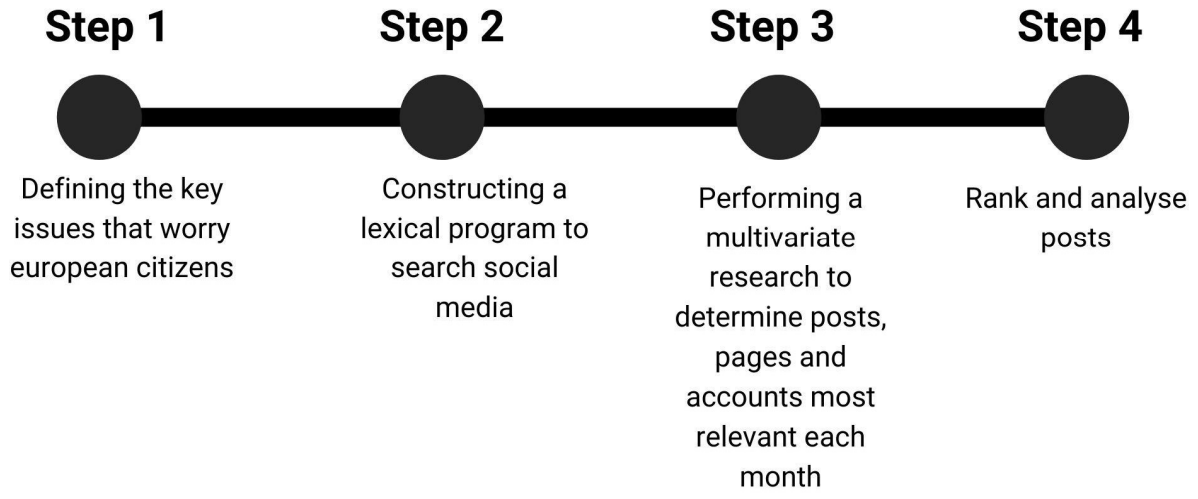


Figure 1: Step by step of methodological framework.

It 's important to stress that the research will use existing materials developed without the researcher’s influence - extant data - and there's no direct contact with individual participants (Salmons, 2018).

The use of digital methods as a method to study and understand the platformization of news is combined with an approach focused on grounded theory based on an inductive process of data collection, aiming to generate theory based on the collection and systematic analysis of data (Strauss & Glaser, 1967).

What tools are we going to use?

CrowdTangle

The entire data collection process on Facebook is performed exclusively through CrowdTangle, a public insights tool owned by Facebook that operates via the available public Graph API[1]. CrowdTangle uses only publicly available data and exclusively tracks public content. Data is downloaded from Facebook pages/groups, which are public entities (CrowdTangle, 2021). We abide by the terms, conditions, and privacy policies of Facebook. We have no access to information about the users who reacted/commented to Facebook content on public pages/groups. For each public post, we have the numeric ID and the name associated to the publishing account, the message contained within the post, the date and time in which the post was initially published, the type of post (link, photo, video etc.), the link attached to the post, the post ID, the "story" description associated to the post, the aggregated number of likes, comments and shares, and the numeric ID associated to the page in which the post is published. Moreover, users' *reactions* include the number of reactions each post got ("angry", "hah", "like", "sad", "wow"). We abide by the terms, conditions, and privacy policies of Facebook

Brandwatch

The entire data collection process on Twitter is performed using a Brandwatch account (owned and operated by Iscte-Iul). Brandwatch[2] is a commercial information retrieving company that operates exclusively within the framework of the Twitter API[3], which is publicly available. We use only publicly available data. Users with privacy restrictions are not included in our dataset. Data is downloaded from Twitter accounts that are public entities. The tools provided by Brandwatch (Brandwatch, 2021) abide by the terms, conditions, and privacy policies of Twitter. Data will include public tweets made on the timelines of public users that correspond to a search query, as well as the number of retweets, replies and mentions corresponding to those tweets. No personal data from the users will be downloaded other than that which is publicly available though the API. We abide by the terms, conditions, and privacy policies of Twitter.

YouTube Data Tools

---

[1] https://developers.facebook.com/docs/graph-api/ and https://developers.facebook.com/docs/graph-api/reference/v2.10/comment

[2] https://www.brandwatch.com/blog/brandwatch-and-the-gdpr-what-you-need-to-know/

[3] https://developer.twitter.com/en/docs/api-reference-index

The entire data collection process on YouTube is performed using the YouTube Data Tools (Rieder, 2015) which are publicly available and operate exclusively by means of the YouTube Data V3 API[4], which is also publicly available. We used only publicly available data. Users with privacy restrictions are not included in our dataset. Data is downloaded from YouTube channels that are public entities. We abide by the terms, conditions, and privacy policies of YouTube. Data will include all data relative to the published public videos made on public channels corresponding to a search query, as well as the number of views, likes, dislikes, favorites and comments corresponding to those videos. This would include the timestamp of the video, it's title and caption and tags. No personal data from the users will be downloaded other than that which is publicly available though the API.

Step by Step

3.1 Step 1: Defining the key issues that worry european citizens

The two main concepts behind EUMEPLAT are those of Platformization and Europeanisation.

Within work package 2[5] we are focusing on the platformization of news, expressed by the way certain key preoccupations of European citizens are displayed on major social media platforms (Facebook, Twitter and YouTube), either by the news media or by other actors on such platforms.

To that end, we referred to the Summer 2020 edition of the Europe-wide survey Eurobarometer (Eurobarometer, 2020), carried out in 34 countries in Europe, to determine the key issues that concerned Europeans. The survey addresses topics such as the political and economic situation in Europe, how Europeans perceive their political institutions, attitudes to European citizenship and other key policy areas. When asked to elect the two most important issues that concern them (see Figure 2), European citizens highlight issues regarding economy and finances (positions 1 and 2), immigration (position 3), health (posicion 4) and environment and climate (position 5).

[4] https://developers.google.com/youtube/v3/getting-started

[5] for a deeper description see Project

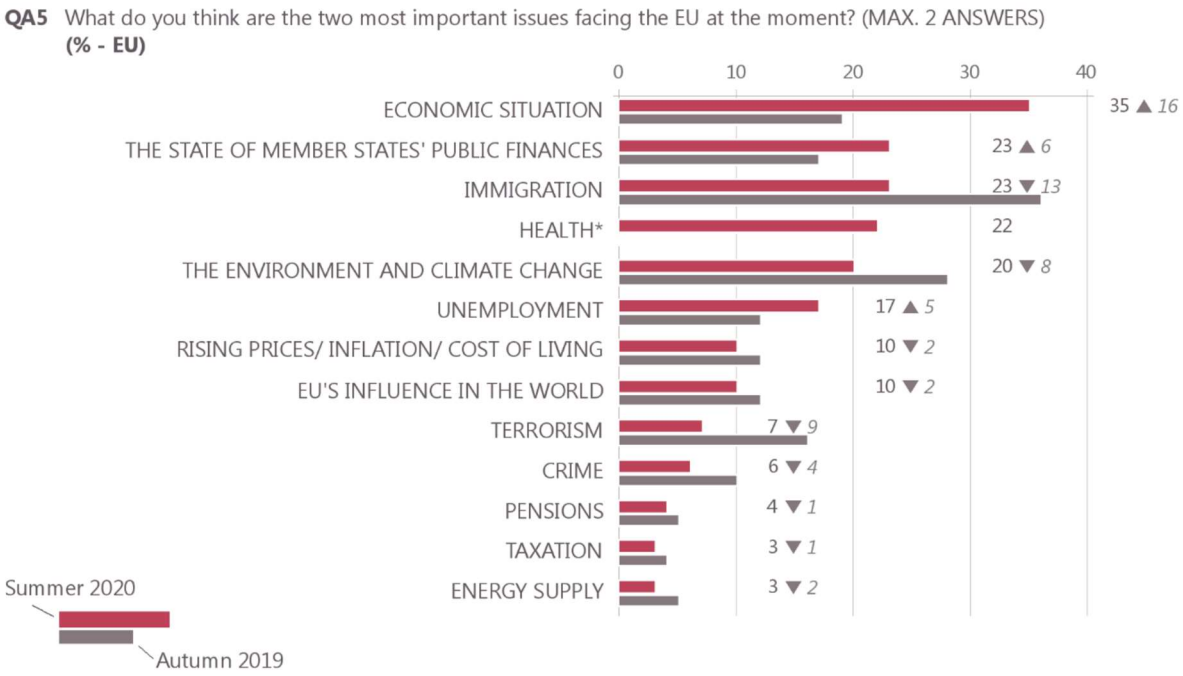


Figure 2: The two main concerns of European. Source: Eurobarometer (2020).

Also, when we look at the survey results at the national level (see Figure 3), we notice some variations in the weight of each issue in each country, but the five referred issues are dominant in most of the countries.

QA5 What do you think are the two most important issues facing the EU at the moment?
(MAX. 2 ANSWERS)
(%)

	Economic situation	The state of Member States' public finances	Immigration	Health	The environment and climate change	Unemployment	Rising prices/ inflation/ cost of living	EU's influence in the world	Terrorism	Crime	Pensions	Taxation	Energy supply
EU27	35	23	23	22	20	17	10	10	7	6	4	3	3
BE	33	20	23	31	26	14	9	8	5	7	5	3	3
BG	36	13	33	23	12	11	16	10	12	8	3	2	4
CZ	26	28	38	14	11	6	17	10	15	8	4	3	1
DK	42	20	24	14	44	19	3	10	5	3	1	1	2
DE	35	32	27	10	24	12	10	15	6	8	2	2	2
EE	37	34	40	12	22	8	6	18	9	5	1	2	3
IE	47	31	15	26	35	13	9	11	4	1	0	3	2
EL	34	23	38	30	6	18	6	14	7	9	2	2	2
ES	43	15	13	37	9	30	5	4	2	1	6	2	1
FR	33	17	21	21	28	18	11	9	8	9	5	2	3
HR	33	25	25	31	11	14	10	10	10	9	3	3	3
IT	42	21	21	24	11	28	11	5	4	4	5	8	3
CY	32	12	53	18	8	16	5	6	15	19	1	2	0
LV	30	23	36	12	15	11	10	8	13	7	3	6	1
LT	25	15	30	16	16	8	12	12	19	9	2	4	2
LU	33	20	21	30	36	16	5	18	1	6	2	2	2
HU	29	22	40	15	20	9	12	10	12	7	4	1	4
MT	22	7	61	30	12	9	7	3	4	8	3	2	2
NL	41	37	20	27	36	9	3	15	2	4	1	1	1
AT	36	21	18	30	21	17	13	7	5	9	6	4	3
PL	21	19	24	20	19	5	18	10	14	6	5	6	6
PT	38	34	10	45	4	28	5	5	5	5	2	2	0
RO	25	17	19	31	14	12	15	9	11	11	6	5	5
SI	33	15	33	37	10	12	5	10	7	7	3	3	2
SK	26	31	35	10	13	11	18	10	14	12	4	2	2
FI	37	44	25	13	30	8	7	14	8	6	1	3	2
SE	38	22	25	14	49	15	4	12	5	7	1	1	5

1st MOST FREQUENTLY MENTIONED ITEM | 2nd MOST FREQUENTLY MENTIONED ITEM | 3rd MOST FREQUENTLY MENTIONED ITEM

Figure 3: The two most important issues the European Union is facing - National. Source: Eurobarometer (2020).

Thanking into account this results we have arrived to four major dimensions[6] of analysis regarding Europe or related to Europe, in the europeanization context of this research program that we've already discussed:

- **Europe**;
- **Health**, as it relates to **Europe**;

[6] Note that although immigration was also one of the most relevant preoccupations, such subject is addressed on WP4.

- **Economic situation**, as it relates to **Europe**;
- **Environment,** as it relates to **Europe.**

3.2 Step 2: Constructing a lexical program to search social media

Query design

Following Rogers (2017) recommendations on query design strategy, it's necessary for us, before query design, to define what may be considered a keyword concerning the four major dimensions we are researching. The set of keywords, locally adapted to the different countries' lexicon and culture, will compose our lexical program. To reach that goal, we started with a few words related to the dimensions of analysis in what contexts those words are being used, with which frequency and by whom. Then, we searched for examples of the use of those keywords in various contexts on the internet and social media - in Google searches, Facebook and Twitter - and collected the words or concepts that were most frequently associated with those words (using affordances provided by the platforms or the tools we are using to study them). You can see examples of that process of keyword search and query construction on the annex 1 to this document.

The idea was to determine what may constitute keywords in our subject taking into consideration: 1) the keyword definition as "'a word which acts as the key to a cipher or code" (Stevenson & Lindberg, 2015); 2) the sense that "the available and developing meanings of known words" (Williams, 1975, p.13); and 3) "the explicit but as often implicit connections which people are making" (Williams, 1975, p.13).

2.2. Task 1: Constructing the lexical programme

To obtain the necessary datasets from CrowdTangle, Brandwatch and YouTube Data Tools[7] (tools we are using for data extraction), we need to define search queries for thata data. The construction of the query is a meaningful and determinant step of the subsequent research, because the query will generate a specific and unique dataset upon which the analysis will dwell and from which results will derive. Therefore, changing the query (remove, change or add one or more keywords) will create a different dataset and different results[8].

[7] For more detail on these tools please refer to section 4

[8] See example of keywords exploration and query construction on annex 1.

In this case, the queries will be composed by a set of keywords specific to each one of the four dimensions of the research: (1) health, (2) economy, and (3) the environment, all combined with (4) Europe.

After each query is established, it will be inserted into the Crowdtangle, Brandwatch or YouTube Data Tools corresponding field to collect the data corresponding to that query and a given timeframe. In some cases, that query will be filtered by a given list of news media using the corresponding platform (please refer to task 2). In this case, the query is the same, but the results - because they are filtered - will be different. From each of these queries (and filters) a specific dataset (in csv format) will be created, each dataset corresponding to each dimension and each extraction.

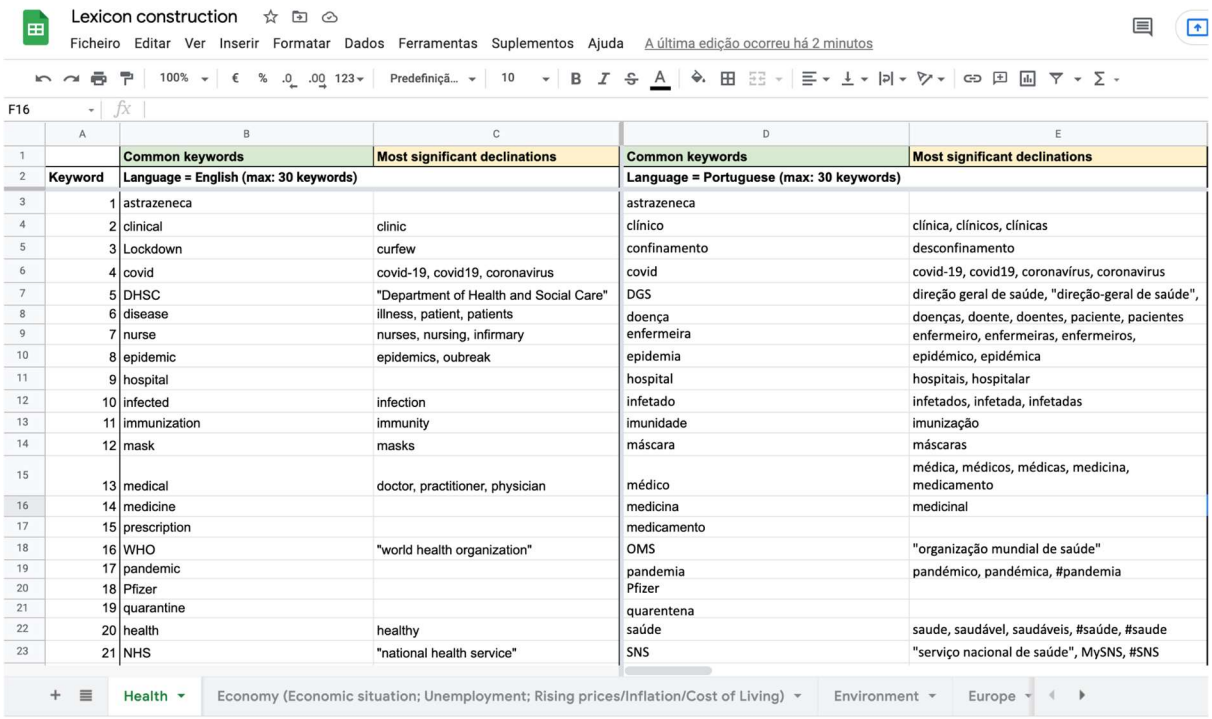
For the development of the queries to use in this research, we started by creating a list of keywords in three parts:

A. **Common keywords:** Keywords that will be common to all the 10 language queries, for each dimension, so as to give a degree of comparability among different countries. These keywords were firstly collected via social media through a “grounded” approach using snowball sampling.
B. **Other suggestions for common keywords:** Keywords that each country felt should be common to all countries but were not in the above group. This was a section open for contributions from each country. After each country's suggestions, we validated and eventually incorporated those keywords into the common group.
C. **Keywords related to national debate:** Keywords that are specific to each country, but relevant enough in the social media debate so as to be included in the search query. Examples: “EstamosOn” (a slogan used by portuguese health authorities on social media) or #Fiqueemcasa (hashtag that translates as #stayathome).

The goal of this approach was to allow each country/language in the EUMEPLAT project to adapt the most relevant keywords to its language (A) and, at the same time, suggest keywords that could be relevant also for the ensemble of the other countries (B). The third category was reserved for keywords considered relevant in one country/language (C) but not necessarily in the other countries/languages.

All the keywords had to be inscribed in the country specific language including the most relevant declinations. Declinations are important because languages vary in the alternativa words they offer to refer to the same issue. For certain countries, more than one language was used, to safeguard the fact that the content to research could be expressed in more than one language. It is the case of Belgium, where French and Flemish were employed, and the case of Spain with Spanish and Catalan.

To construct the queries to use in this research, we invited each Partner to fill in the Lexicon construction document with A, B and C keywords for each one of the four dimensions. There was a specific sheet for each dimension: (1) Health; (2) Economy; (3) Environment; (4) Europe. On Figure 4 you can see the look and feel of the document.



Lexicon construction

Ficheiro Editar Ver Inserir Formatar Dados Ferramentas Suplementos Ajuda A última edição ocorreu há 2 minutos

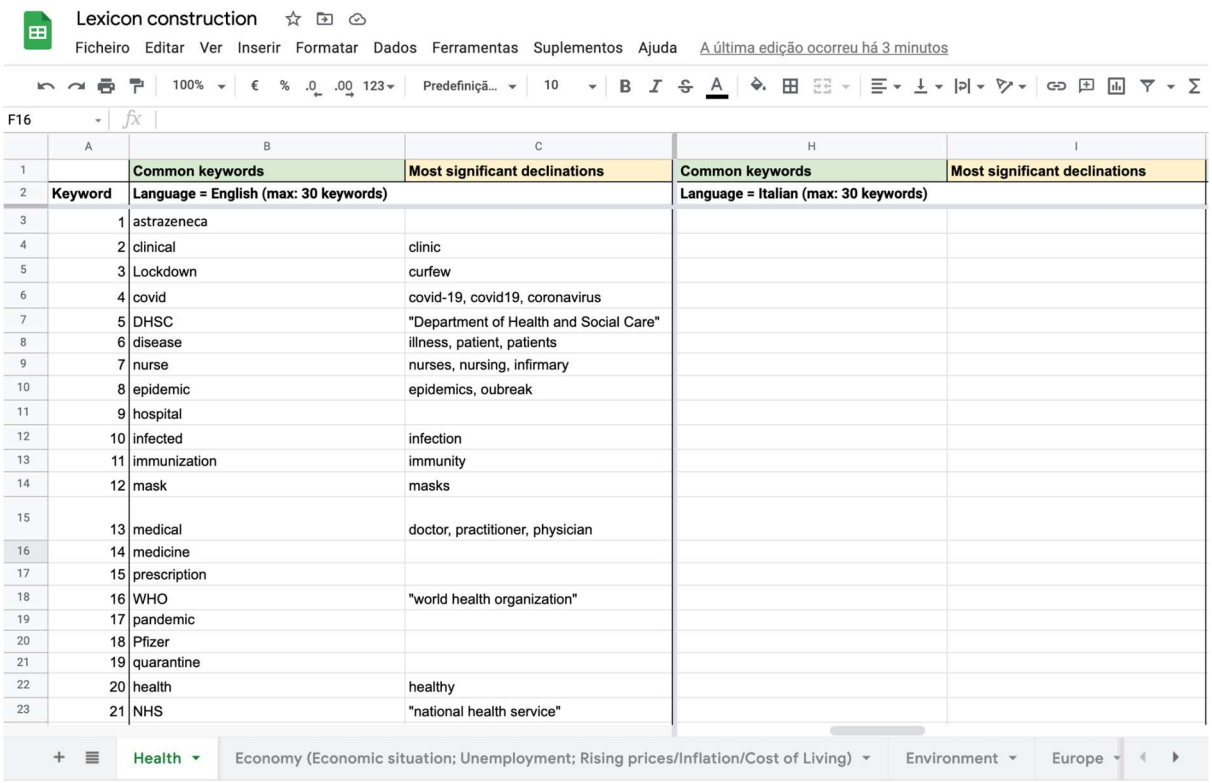
		Common keywords	Most significant declinations	Common keywords	Most significant declinations
Keyword		Language = English (max: 30 keywords)		Language = Portuguese (max: 30 keywords)	
1	1	astrazeneca		astrazeneca	
2	2	clinical	clinic	clínico	clínica, clínicos, clínicas
3	3	Lockdown	curfew	confinamento	desconfinamento
4	4	covid	covid-19, covid19, coronavirus	covid	covid-19, covid19, coronavírus, coronavirus
5	5	DHSC	"Department of Health and Social Care"	DGS	direção geral de saúde, "direção-geral de saúde",
6	6	disease	illness, patient, patients	doença	doenças, doente, doentes, paciente, pacientes
7	7	nurse	nurses, nursing, infirmary	enfermeira	enfermeiro, enfermeiras, enfermeiros,
8	8	epidemic	epidemics, oubreak	epidemia	epidémico, epidémica
9	9	hospital		hospital	hospitais, hospitalar
10	10	infected	infection	infetado	infetados, infetada, infetadas
11	11	immunization	immunity	imunidade	imunização
12	12	mask	masks	máscara	máscaras
13	13	medical	doctor, practitioner, physician	médico	médica, médicos, médicas, medicina, medicamento
14	14	medicine		medicina	medicinal
15	15	prescription		medicamento	
16	16	WHO	"world health organization"	OMS	"organização mundial de saúde"
17	17	pandemic		pandemia	pandémico, pandémica, #pandemia
18	18	Pfizer		Pfizer	
19	19	quarantine		quarentena	
20	20	health	healthy	saúde	saude, saudável, saudáveis, #saúde, #saude
21	21	NHS	"national health service"	SNS	"serviço nacional de saúde", MySNS, #SNS

Health | Economy (Economic situation; Unemployment; Rising prices/Inflation/Cost of Living) | Environment | Europe

Figure 4: Screen Capture of the lexicon construction document (displaying part of the keywords used in the "Health" dimension in English and Portuguese).

There were two fixed columns guiding the filling of the document, one with the order and count of the keywords and other with the keyword (and its declinations) in English. This was meant to serve as guidance for partners to fill the keywords in their own language (see Figure 5). For each language, there are also two columns (Figure 2), one for the keyword and other for the declinations of that keyword.

Partners fulfilling this lexicon were instructed to translate the keywords into their own language not as a literal translation but as a free translation, taking into consideration the word that is most commonly used to express that keyword in their language. This is because what this framework is aimed at capturing is not the literal use of this list of keywords in each country but rather the discourse around those keywords in each language, both by the media and by non-professional users. Also, for the same reason, when fulfilling the possible declinations of a given keyword, partners were asked to fulfill not all the possible declinations (which, in some languages, would be numerous) but only the most relevant declinations (as in "used").



		Common keywords	Most significant declinations	Common keywords	Most significant declinations
Keyword		Language = English (max: 30 keywords)		Language = Italian (max: 30 keywords)	
1	astrazeneca				
2	clinical	clinic			
3	Lockdown	curfew			
4	covid	covid-19, covid19, coronavirus			
5	DHSC	"Department of Health and Social Care"			
6	disease	illness, patient, patients			
7	nurse	nurses, nursing, infirmary			
8	epidemic	epidemics, oubreak			
9	hospital				
10	infected	infection			
11	immunization	immunity			
12	mask	masks			
13	medical	doctor, practitioner, physician			
14	medicine				
15	prescription				
16	WHO	"world health organization"			
17	pandemic				
18	Pfizer				
19	quarantine				
20	health	healthy			
21	NHS	"national health service"			

Figure 5: Screen capture of lexicon construction document (displaying the "Health" dimension). Fixed columns (English) + language specific columns (Italian).

Step by step procedure for each country

For each sheet that corresponds to each dimension these were the steps adopted by each partner in the procedure to construct a set of queries similar for all countries:

1. Select the pair of columns reserved for the partner's language;
2. Fill in common keywords and for each one indicate on the adjacent cell the most significant declinations. Again: this list of keywords could consist in local adaptations of generic words. We did not seek literal translations but rather adaptations of the words or concepts listed in each of the 10 EUMEPLAT languages. Also, not declinations were to be inscribed; only those considered most relevant regarding their use on social media;
3. If necessary, fill in the second field with other suggestions for common keywords and indicate, for each one on the adjacent cells, the most significant declinations;
4. Fill in keywords related to national debate and for each one indicate on the adjacent cell the most significant declinations.

How to determine these 4 sets of keywords?

To determine the keywords that compose the "common keywords" set, we performed several

search queries on Google Search, Google Trends, Facebook and Twitter, using a snowballing approach. The goal was to identify the keywords that were most frequently used on search and social media, when users of social media platforms referred to the 4 dimensions of the study (Health, Economy, Environment and Europe) over the last 12 months.

For those 4 dimensions, we used 2 or 3 basic keywords and analysed which related keywords were most frequently used. Some of those related keywords were then also used as initial search keywords, thus generating a snowball effect. The initial keywords used for each theme were as follows:

- Health: “health”; “covid”; “vaccine”
- Economy: “economy”; “unemployment”; “rising prices/inflation”
- Environment: “environment”; “climate change”; “global warming”
- Europe: “europe”; “EU”

Following, there are some examples of an extraction of roughly the same query (Health crossed with Europe) for 4 different countries (Portugal, Spain, Italy, Czech Republic) over 7 days (see Figure 6 to 9 respectively), on Facebook pages. We display only a short selection of some of the data available (in these screenshots: page name; page category; total interactions; part of the message content).

Page Name	Page Category	Total Interactions	Message
Camilo Lourenço	ACTIVITY_GENERAL	11534	Antes da ordem do dia: Ainda o MEL e a denúncia feita pelo Dia D👍O estudo
Público	MEDIA_NEWS_COMPANY	2516	🔴 Autoridades de saúde britânicas contactam adeptos do Chelsea e do City q
André Ventura	POLITICIAN	2495	Mesmo com o aparente recuo de Espanha, estamos a ser o saco de pancada d
Francisco Moita Flores	MOVIE_WRITER	2038	O MAPA DO SUBDESENVOLVIMENTO E DA INJUSTIÇA: Fez bem a Direção
Ser super mãe é uma treta	DIGITAL_CREATOR	1901	É um bom resumo.
André Ventura	POLITICIAN	1275	O CHEGA terá, nos dias 2 e 3 de Julho, no Algarve, um Conselho Nacional de
Luís Costa - paraciclista	ATHLETE	1210	Vice-campeão da Europa em contrarrelógio. Pai, esta é para ti. Obrigado à UVI
Habeas Corpus	NON_PROFIT	1166	A IDEIA DE INJECTAR CRIANÇAS SEM A AUTORIZAÇÃO DOS PAIS OU A F
A BOLA	TOPIC_NEWSPAPER	984	#abola 📰 Capa do dia 1️⃣ 3 de Junho Mais uma no cesto 🏀, Sporting sagra-s
Daniel Oliveira	JOURNALIST	836	Ryanair: ser capacho não é estratégia económica A Ryanair recebe apoios do I
O JOGO	MEDIA_NEWS_COMPANY	765	Primeira página de 08/06/2021 Ler O JOGO está à distância de 1 clique! : http:/
Francisco Louçã	PERSON	744	Se fosse só parolice futebolística Não correu “na perfeição”, diz o primeiro-mini
Público	MEDIA_NEWS_COMPANY	670	Uma comissão parlamentar sueca publicou esta quinta-feira uma revisão sobre
Cristóvão Norte	POLITICIAN	569	👇Entrevista BBC Não tenho o direito de ser cúmplice silencioso da grave intox
Raquel Varela	PERSON	523	Em Estocolomo hoje estão 27 graus, acabei de ouvir na Antena 1. O calor no N

Figure 6: Example relative to Portugal.

Page Name	Page Category	Total Interactions	Message
La Vanguardia	NEWS_SITE	9729	Dos de la madrugada a bordo del tren que une Bombay y Goa. Cristina I
Juanma Moreno	POLITICIAN	4236	Sumamos una muy buena noticia para la Sanidad Pública andaluza: rec
Pedro Sánchez Pérez-Castejón	POLITICIAN	3169	Hoy hemos asistido a la entrega de la medalla conmemorativa del 250 a
PEKE SPIIBERG	TOPIC_JUST_FOR_FUN	3165	He oido cuando hablan tonterias de planes de electrificación del transpo
VOX España	POLITICAL_PARTY	1737	❌ VOX no va a apoyar ningún "certificado Covid" que privilegie a los va
Ignacio Escolar	ACTIVITY_GENERAL	1720	España ha liderado la vacunación de los mayores en Europa, con una re
Yo SOY ROJO	NEWS_SITE	1443	Casado no sabe qué hacer ya...
elDiario.es	NEWS_SITE	1433	‼ Sanidad estudia priorizar en la vacunación frente a la COVID-19 la Sel
Boticaria Garcia	ACTIVITY_GENERAL	1227	La pregunta de la semana: ¿Las vacunas llevan imanes? 🧲 1️⃣ Para er
elDiario.es	NEWS_SITE	1057	La reticencia vacunal ha estado en España por debajo de las previsione
Lucía, mi pediatra	PERSONAL_BLOG	949	Al día se suicidan en España diez personas, una de ellas es un adolesce
Murciasalud	GOVERNMENT_ORGANIZATION	905	El #EquipoECMOArrixaca logra mantener con vida a un recién nacido de
Las huellas del pasado	SOCIETY_SITE	891	La historia de Venecia comienza alrededor de 400 d.C. Las primeras per
Adriana Lastra	POLITICIAN	853	A pesar de una derecha que se ha dedicado a obstaculizar los esfuerzos
La Vanguardia	NEWS_SITE	853	🔴🔴 ÚLTIMA HORA. El expresident recurrió al alto tribunal europeo la
Informativos Telecinco.com	NEWS_SITE	759	Un bebé de cinco meses se ha convertido en el primer paciente europec
PP Andaluz	POLITICAL_PARTY	739	💚 Andalucía, pionera: primera región europea en contar con certificado

Figure 7: Example relative to Spain.

Page Name	Page Category	Total Interactions	Message
Andrea Scanzi	MOVIE_WRITER	62375	"L'ultima che ho sentito dalla Meloni è stato quando Draghi e Speranza hanno t
Giuseppe Conte	POLITICIAN	52473	**LA MIA INTERVISTA ODIERNA AL CORRIERE DELLA SERA** *Giuseppe Conte a tutto*
Giuseppe Conte	POLITICIAN	47716	La regione Calabria andrà al voto e questo appuntamento merita il massimo im
Leonardo Cecchi	POLITICIAN	16547	"L'ultima che ho sentito dalla Meloni è stato quando Draghi e Speranza hanno t
SportMediaset	NEWS_SITE	7234	Bufera in Spagna dopo la positività al Covid-19 di Busquets ‼ ➡ https://bit.ly/3
Maria Giovanna Maglie	JOURNALIST	6348	Ieri Tienanmen, oggi Hong Kong Oggi il covid. Tutti mostri cinesi. La dittatura co
Virginia Raggi	POLITICIAN	6162	Gli Street World Championships 2021, uno dei più importanti appuntamenti inte
MoVimento 5 Stelle	POLITICAL_ORGANIZ	4410	ABBIAMO LAVORATO PER LA TENUTA DEL PAESE DURANTE LE FASI PIÙ
Il Casciavait	ACTIVITY_GENERAL	4251	Al 2 giugno la situazione è la seguente: - Maignan preso ed ufficializzato dopo
Pillole di Ottimismo	SCIENTIST	3771	NUMERI IN PILLOLE - 6 giugno 2021 - Paolo Spada Buonasera e ben ritrovati
Utopia Reale	ACTIVITY_GENERAL	3502	Ammettendo anche che quelle stabilite dal #Governo #Draghi alla fine di aprile
Stefano Bonaccini	POLITICIAN	3268	Ieri in Emilia-Romagna 45mila vaccinazioni. Forza, ripartiamo come Paese 🇮🇹
Utopia Reale	ACTIVITY_GENERAL	3233	Quando va indossata la #mascherina? Sempre, risponderebbe chiunque stand
Fabio Massimo Castaldo	POLITICIAN	2790	ANCHE LE REGIONI RIVOGLIONO IL VITALIZIO! Dopo lo scandalo del vitalizi
Report	TV_SHOW	2664	#Report #DaRivedere "Splendori e miserie dei signori del calcio" di Daniele Aut
ANSA.it	MEDIA_NEWS_COMP	2399	Raggiunta quota 600 mila vaccinazioni in un giorno in Italia. L'Italia è al second
Utopia Reale	ACTIVITY_GENERAL	2348	#Mascherine: il tratto probabilmente più distintivo della #pandemia. Uno strume

Figure 8: Example relative to Italy.

Page Name	Page Category	Total Interactions	Message
Tomio Okamura - SPD	POLITICIAN	1409	Můj dnešní rozhovor pro Parlamentní listy - hlavní programové rozdíly mezi politickými st
ČT24	TV_CHANNEL	461	Jsou očkovaní, úlevy spojené s rozvolněním pro ně ale neplatí. Většina Čechů žijících, p
ODS - Občanská demokratická str	POLITICAL_PARTY	227	„Vážené poslankyně, vážení poslanci, dnešek je důležitý den pro českou demokracii. Po
Události Ostrava	TV_SHOW	172	Možnost registrovat se k očkování proti covidu-19 se dnes v noci otevřela i mladým liden
CNN Prima NEWS	MEDIA_NEWS_COMPANY	165	Kde najít svůj QR kód a k čemu všemu slouží?
Barbora Kořanová	POLITICIAN	128	PROČ DNES V POSLANECKÉ SNĚMOVNĚ PODPOŘÍM VLÁDU? Zdravím vás z Posla
Evropská komise v ČR	POLITICAL_ORGANIZATION	127	Česko cz je společně s Chorvatskem, Řeckem, Polskem, Německem, Dánskem a Bulhar
Události Ostrava	TV_SHOW	61	Očkovaní si už v Česku mohou zajít třeba do bazénu nebo na pivo bez povinného testu,
EVROPA 2	RADIO_STATION	45	Co na obsazení říkáte? 😂👇 #film #music
CNN Prima NEWS	MEDIA_NEWS_COMPANY	41	Má evropský covid pas smysl?
Ministerstvo práce a sociálních vě	GOVERNMENT_ORGANIZATION	40	Zveme vás na diskusi mezi odborníky v oblasti péče o ohrožené děti, která se bude kona
Zpravodajství FTV Prima	ACTIVITY_GENERAL	35	Kde najít svůj QR kód a k čemu všemu slouží?
Luděk Niedermayer	POLITICIAN	31	Další očkovací update Země EU/EEA již obdržely přibližně 300 milionů dávek vakcín. Blí
TOP STAR	ACTIVITY_GENERAL	22	Vše, co potřebujete vědět o covidpasu.
iROZHLAS.cz	MEDIA_NEWS_COMPANY	20	Oslovení odborníci očkování dětí vítají, zároveň je podle nich nutné, aby se dál nechaly
Vlastimil Válek	POLITICIAN	19	!!PRAVDA A LŽI O COVID PASU!! Je potřeba zdůraznit, že: 👉 Certifikát bude k dispozi
Partička	TV_SHOW	19	Vše, co potřebujete vědět o covidpasu.
Zpravodajství FTV Prima	ACTIVITY_GENERAL	16	Má evropský covid pas smysl?
Události Olomouckého kraje	TV_SHOW	12	Možnost registrovat se k očkování proti covidu-19 se dnes v noci otevřela i mladým liden

Figure 9: Example relative to Czech Republic.

2.2. Task 2: Establishing a list of news media

Taking into account that one of the goals of EUMEPLAT is to analyse the platformization of news in Europe, research into how the news media treat the four selected research dimensions on social media is necessary. To this end, we will perform a social media data extraction using the same set of keywords used on social media, but directed at mainstream news media publications on social media platforms.

To do so, we invited our partners to establish a list of up to the 30 most relevant news media outlets present on social media platforms in their country. The goal is to research how those news media outlets express their journalism on social media on the four dimensions of the study. This approach, of course, aims to investigate in what way news is “platformized” around the issues at study and compare that “platformization” with the way non-professional users express on those same platforms.

How do we define news media?

In general, news media are those mass media that provide news coverage for the general public or a target public. Before the Internet, news media were what we now call legacy media (print, radio, and TV). More recently, online newspapers which follow and reproduce the same practices, organization, legal, ethical, and deontological frameworks fall also into the concept of news media.

Recognised media outlets should have mandatory registration as “media” and be supervised by a regulatory governmental agency (where available) and have ‘journalistic activity’ produced

by ‘journalists’. The regulation aims to ensure civil and penal responsibility of the media. The broad legal framework within which the media operate derives in the first instance from international law.

In many EU member States, the material scope of the media regulatory framework is limited to audiovisual media services as defined by the AVMS Directive, but in others specific media laws establish administrative obligations, such as entering a public register or some form of content regulation.

The news media production must fit into the normative understanding of journalistic practice and media integrity, and media companies are required to hire only licensed journalists (with a press card, where available).

Media integrity refers to the ability of a news media outlet to serve the public interest and democratic process, making it resilient to institutional corruption within the media system, economy of influence, conflicting dependence and political clientelism. Media integrity encompasses the following qualities of a media outlet: independence from private or political interests; transparency about own financial interests; commitment to journalism ethics and standards; responsiveness to citizens. This media transparency reflects the relationship between civilization and journalists, news sources and government.

The exercise of journalism requires prior validation and certification. Journalism practice is obliged to fit into a very detailed and strict legal framework, involving several evolutionary steps: full compulsory education (different from country to country), peer evaluation in a newsroom (as a trainee for up to two years), production of news stories published in recognised media outlets (also with mandatory registration and supervised by a regulatory governmental agency) and, finally, through an administrative process where each candidate has to prove to meet all the requirements necessary to receive a press card, the letter constituting a prerequisite to work as a journalist. In several countries, the press card is dependent on peer evaluation by the administrative body or professional orders, like professionals such as lawyers, medical doctors, architects, etc.

What to Include?

This part of the project aims to understand how mainstream news outlets cover Europe and European issues, contributing to form a certain type of public opinion vis à vis those topics. As such, we will only be concentrating on non-alternative online news media that fit into the above definition and that cover a broad range of topics. For the purposes of this study, we are *not* focusing on news media that may fall within the ‘niche’ publishing market. News agencies may also be included within the scope of online news outlets, particularly due to the fact that many of their news pieces end up being assimilated by the mainstream press, both offline and online.

The objective then would be to include the most important media in any particular country, on the basis of their presence in terms of numbers of followers on social media. Although there may be some discrepancy, in some cases, between the analog and digital versions of news outlets in terms of audience numbers, we have chosen to define importance as corresponding to the number of online followers on the platforms that we are studying. This is because the online mainstream news outlets on the basis of which we are performing our selection comprise news media that are only digital as well as those that are both analog and digital. In our opinion, it only makes sense to focus on the digital as a common element among these news media.

Furthermore, our objective is that of making the selection of online news outlets - encompassing press, TV and radio - only on the basis of their impact on social media (i.e. in terms of number of followers), rather than on their political slant (i.e. pro-EU and anti-EU; left and right-wing; populist and bourgeois). This is because we want to understand the most pressing concerns that are articulated on those digital outlets that have the greatest influence in terms of public opinion formation in the time-frame that we are studying. The bias of the news outlets will not, indeed, be taken into account at this stage of the project. However, it may indeed be an important factor to consider when analysing the data arising from this research. Analysis of the data itself will comprise another stage of WP2.

For that, we asked each partner country to deliver a sequentially-ordered list with a maximum of 30 mainstream online news media with the highest number of followers on both Facebook and Twitter as a total sum (in case any particular medium was represented on only one of these platforms, it could be included, as long as its following was within the top 30) . This allows us to place all countries on an equal footing, which is important for the purposes of comparability, whilst simultaneously introducing a ranking factor, which allows us to account for the different media landscapes for different countries.

Countries, such as Belgium and Spain, which have more than one language, were asked to include the dominant and second most dominant languages. This second language may require justification, taking into account that other languages may be present as well in that national context. Nevertheless, the criteria discussed pointed to the possibility of including a second language of a region in which a strong separatist movement existed, which had been publicly legitimated in the political public sphere. As such, our Spanish partners included Castilian and Catalan, while our Belgian partners included Flemish and French.

Considering what is contextualized above, this is what we asked partners to do when establishing the news media lists:

1. Select the most important online news media (press, TV, Radio or news agency) in each country/language, that have presence on social media (Facebook and Twitter simultaneously). We consider as news media, the digital mainstream media that produce news directed towards a general audience, covering a broad range of topics. Such mainstream

generalist online news outlets (such as TV stations**,** Radio stations, general online news sites or news agencies).

2. Order those news media by their following on social media (Facebook and Twitter as a total sum; in case any particular medium is represented on only one of these platforms, it can be included, as long as its following is within the top 30) . Please do not prioritize the importance of a given news media outlet in the offline world but rather in the online world, particularly on social media platforms;

3. Select up to the 30 most important online news media outlets by the above criteria. If you do not have 30 news outlets, please inscribe those that you have. If you have more than 30, please restrict your list to 30.

4. In the case of multi-language countries, please provide a different list of each language that you will be using (as the query in one language will only work on news media that are written in that language).

The resulting lists of up to 30 social media accounts (Facebook and/or Twitter) of the most relevant mainstream news media for each country (see Figure 10 as an example). Within this framework, these 12 media lists (from 10 countries) will be subjected to the same queries as general platform users in each country, permitting to compare the way the professional news media perform the researched four dimensions on social media in comparison with other non-professional users.

	A	O	P	Q	R
		GERMANY			
		News media	**Facebook handle**		**Twitter handle**
1		Tagesschau	https://www.facebook.com/193081554406	@tagesschau	https://twitter.com/tagesschau
2		DER SPIEGEL	https://www.facebook.com/38246844868	@derspiegel	https://twitter.com/derspiegel
3		Bild	https://www.facebook.com/25604775729	@BILD	https://twitter.com/BILD
4		DW News	https://www.facebook.com/24369314439	@dwnews	https://twitter.com/dwnews
5		ZEIT ONLINE	https://www.facebook.com/37816894428	@zeitonline	https://twitter.com/zeitonline
6		WELT	https://www.facebook.com/97515118114	@welt	https://twitter.com/welt
7		Süddeutsche Zeitung	https://www.facebook.com/215982125159841	@SZ	https://twitter.com/SZ
8		ZDF heute	https://www.facebook.com/112784955679	@ZDFheute	https://twitter.com/ZDFheute
9		stern	https://www.facebook.com/78766664651	@sternde	https://twitter.com/sternde
10		RTL.de	https://www.facebook.com/179133132098813	@RTLde	https://twitter.com/RTLde
11		ntv Nachrichten	https://www.facebook.com/126049165307	@ntvde	https://twitter.com/ntvde
12		RTL Aktuell	https://www.facebook.com/119845424729050	@rtl_aktuell	https://twitter.com/rtl_aktuell
13		FAZ.NET - Frankfurter Allgemeine Zeitung	https://www.facebook.com/346392590975	@faznet	https://twitter.com/faznet
14		FOCUS Online	https://www.facebook.com/37124189409	@FOCUS_TopNews	https://twitter.com/FOCUS_TopNews
15		taz (die tageszeitung)	https://www.facebook.com/171844246207985	@tazgezwitscher	https://twitter.com/tazgezwitscher
16		RT DE	https://www.facebook.com/472061332924101	@de_rt_com	https://twitter.com/de_rt_com
17		Handelsblatt	https://www.facebook.com/104709558232	@handelsblatt	https://twitter.com/handelsblatt
18		Tagesspiegel	https://www.facebook.com/59381221492	@Tagesspiegel	https://twitter.com/Tagesspiegel
19		FOCUS: Gut für Mich	https://www.facebook.com/366193510165011		
20		Berliner Morgenpost	https://www.facebook.com/morgenpost	@morgenpost	https://twitter.com/morgenpost
21		Netzpolitik.org	https://www.facebook.com/netzpolitik	@netzpolitik	https://twitter.com/netzpolitik

Figure 10: Example of part of Germany media list.

Limitations of the method

Using digital methods can be a straightforward choice when researching online digital platforms. But it also has some important methodological caveats that researchers embracing this route should bear in mind, namely taking into account that we will be researching inside given platforms and therefore subject to the affordances and limitations of those platforms:

a) First of all, this framework will be using tools for extracting data from social media platforms (Facebook, Twitter and YouTube) that all operate through the authorized API (Application Programming Interface) for each of those social media platforms. This means that **the data points available for research and analysis are only the ones that are accessible though those API's**. On the other hand, that also means that all those data points are compliant with each platform's terms and comprise only public data, thus ensuring compatibility with GDPR, namely as it relates to data anonymity;
b) For analysis, we will be considering the 10 most relevant posts each month, on each platform, on each of the dimensions that are being researched. That means **we will be considering that the most relevant posts are those that have had the most interactions, reach or relevance (depending on the platform)** in that given month (of all the posts on that dimension, as captured by the query). Interactions, reach or relevance are a proxy for the quantity of social media users that may have contacted with the message contained in each post or publication. But those are metrics developed, owned and controlled by the platform;
c) Furthermore, for each country, we will filter the results by language and/or country/geolocalization, in order to filter out results for each country and allow comparison between countries. However, we have to pay attention to the fact that **each platform's API uses different methods to attribute the geolocalization of a publication or author and that information is not always available**, so some posts may not be considered as their country/geolocalization cannot be determined. Also, some posts/publications may be directed at countries outside Europe, despite being published in Europe. We will have an "off-topic" category to discard those publications (as explained later in this document).
d) Finally, this framework will only provide the analysis with a sample of social media posts in a given time frame. Although significant, that sample does not represent the universe of publications on social media in that time frame or even the totality of social media posts published about the four dimensions of the study. The extraction of posts will limit to those that correspond to query and the analysis will focus only on the 10 posts with most interactions, reach or relevance each month. This means that a large number of posts about these issues are not captured by the query and, of those that are, only the most significant are analysed.

3.3 Step 3: Performing a multivariate research to determine posts, pages and accounts most relevant each month

For this research program, data will be collected from three different social media platforms: Facebook, Twitter and YouTube. For each one of these platforms, specific types of data will be retrieved, corresponding to the queries and relative to the four dimensions that compose the study, in some cases regarding its professional content or non-professional content. Figure 11 specifies what type of data will be collected from each platform.

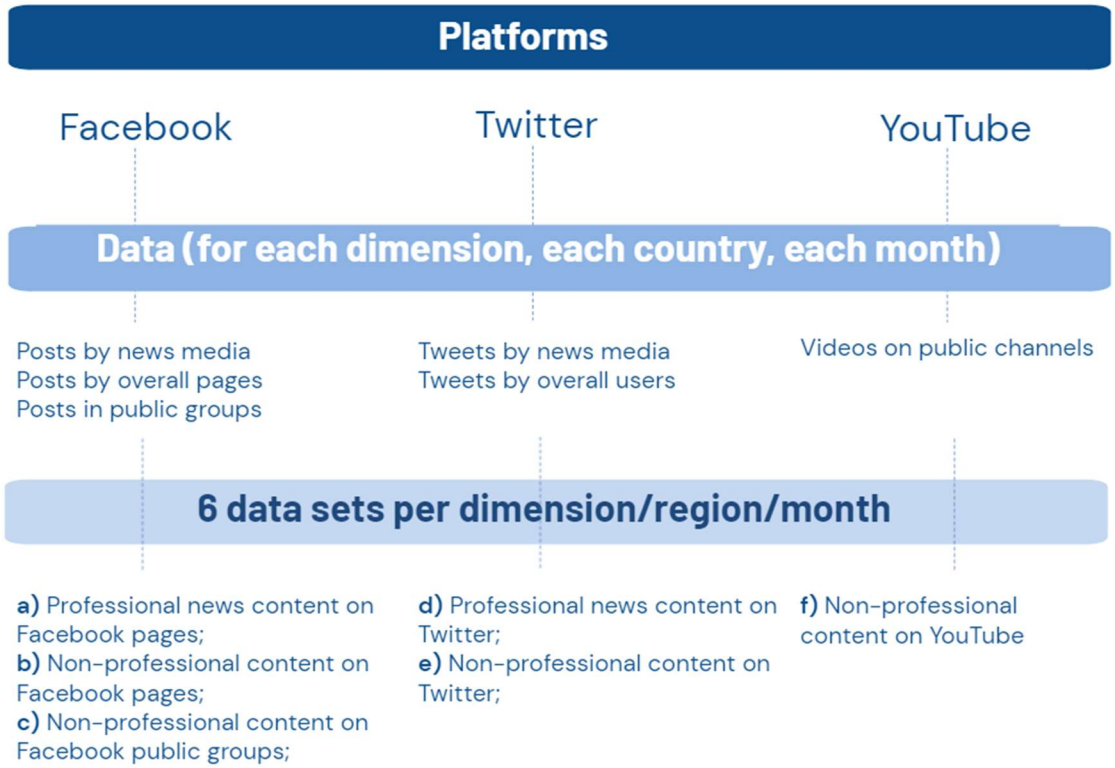


Figure 11 - Diagram specifying the data sets that will be extracted per month and per country from three social media platforms, regarding each dimension, with filtering by professional or non-professional content.

These extractions will form the basis for the sample on which further analysis will operate. We should note that each platform (and the corresponding API) has its own architecture, affordances and data collection and availability. Therefore, the data collected is not exactly the same in different platforms. However, because the queries and the time frame are the same, the research framework is designed precisely to allow some degree of comparison between different platforms, different actors and different countries for the same object (for example: content about Europe by professional media or non-professional users in two different countries.

These extractions will be made in CSV format (or similar) to a shared drive available to all the partners. The goal is to provide partners in this project with a sample of social media posts regarding the dimensions in study that are similar in each country, corresponding to the same query, extracted in the same time frame and filtered by the same professional or non-professional content. This will comprise the raw data from which a significant sample of social media posts will be analysed and categorized.

3.4 Step 4: Rank and analyse posts

After the queries are created, the filters are established and the tools are prepared, the dataset extractions may begin. As stated previously, this data will be extracted from the social media platforms (according to the corresponding query, in each country and filtered by professional/non-professional content) in CSV or similar format. In the extraction process, the data sets will be downloaded with posts for each month already ranked by interactions (Facebook), reach (Twitter) and relevance (YouTube).

According to Facebook, the “interactions” metric corresponds to the sum of all reactions to a post (Like, Love, Care, Haha, Wow, Sad and Angry), all comments on that post and all shares made of it. Interactions is both the overall most important metric in Facebook, also driving the ranking algorithm and therefore the popularity of posts and pages, and the default metric of the tool we are using to extract data from Facebook (Crowdtangle). In this way, our extraction will rank those pieces of content that, for a given query, gathered the most interactions (as a proxy for attention) in a given time frame.
In the case of Twitter, reach is the most recent metric used to track the popularity of content items. According to Twitter, reach corresponds to the number of people estimated to have seen a given post. This calculation takes into account metrics such as followers, engagement, page ranks and estimated views of a given piece of content. Likewise, this means that the data extracted will be ranked by those publications that, estimatedly, were viewed by the most users.

Finally, in the case of YouTube, the metric of reference is relevante. Extractions to perform using the YouTube Data Tools will display the videos that the platform algorithm considers the most relevant towards a given search query. This means that our results will be approximate to those that a regular user would obtain when performing the same query on YouTube.

A synchronic analysis of social media posts

The goal of this framework is to make way for a synchronic view of the platformization of news in the 10 countries involved in the project. That means the objective is not the evolution of the

phenomenon but its current status in a given time frame in the 10 countries. Therefore, the framework is prepared to extract the same data (that is, corresponding to similar queries) in the 10 countries in exactly the same period.

We foresee three extractions, comprising a month each, to each of the four dimensions, 10 countries and two types of users (professional/non professional). In total (see Figure 12), this will mean six datasets, per extraction and per country, corresponding to a total of up to 240 month posts for analysis. The total for the three months will be up to 720 posts for analysis for each country, up to 7.220 for the 10 countries (8.640 if we consider 12 languages).

Month	Dimensions	Dataset	Post per dataset	SubTotal (country)
1	4	6	10	240
2	4	6	10	240
3	4	6	10	240

Figure 12 - Diagram with the total number of extractions and posts for analysis for each month in each country.

The option for a monthly extraction has to do with the presumed rhythm of information circulation on social media. A monthly timeframe analysis will capture the most important actors or posts in that month, whereas the extended analysis over a período of three month will provide a more complete synchronic view of the state of the platformization of news in those 10 countries, which is the aim of this framework.

As stated before, the aim is to observe the reality of news on social media from a synchronic rather than diachronic perspective. However, the analysis of just one week or of 3 or 4 subsequent weeks in a month would again, presumably, generate a disproportionate weight of the actors and posts that are viral in one given period, thus not offering a fair synchronic view on the platformization of news. To curtail that problem we will perform a monthly extraction in three consecutive months, from September 2021 to November 2021 - with a distinct dataset for each month, in each country and in each dimension. With this approach we will cover the continued relevance span of three months.

The total number of the posts that will be the object of further analysis in the implementation of the framework (see parte 3), are what constitutes our sample. Inasmuch as this sample includes a selection of posts/publications most relevant in four dimensions, three social media platforms and both professional and non-professional content, we believe it will be able to paint a significant picture of the state of the platformization of news in each country. On the other hand, inasmuch as the framework is the same and is similarly implemented in the ten countries, it will be possible to establish significant comparisons between countries, either in reference to dimensions, platforms and/or professional or non.professional users.

Because the framework uses, for each dimension, similar queries, both for different countries and for different platforms, as well as for professional or non-professional content, all the variables will be comparable. Results will of course be able to compare different countries in all or each of the four dimensions, as well as only on professional or non-professional content or, alternatively only on one of the three social media platforms. Likewise it will also be possible to compare professional to non-professional content as well as Facebook to Twitter or to YouTube as well as Europe to Health or Economy or Environment (see Figure 13).

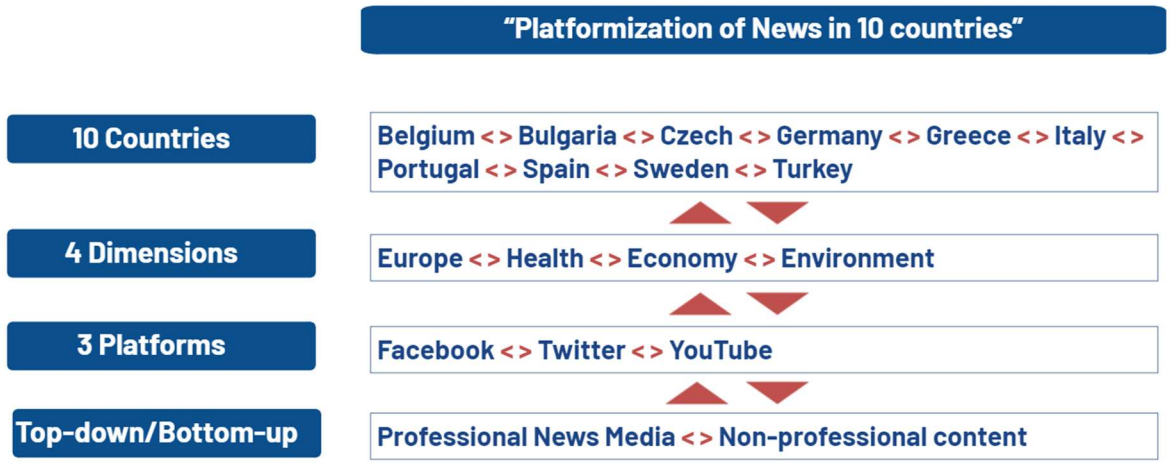


Figure 13 - Diagram showing the possible comparisons between countries, dimensions, platforms and users within the Eumeplat framework for studying the platformization of news.

4 The platformization of news in 10 countries: implementing the framework

The framework described above was designed to study the platformization of news in the 10 countries involved in the EUMEPLAT project in the context of Europe and europeanization. The goal is to collect a significative sample of social media posts, on three of the most relevant social media platforms, and compare the results in different countries, by professional or non-professional news producers and in four different dimensions: Europe; Health when related to Europe; Economy when related to Europe; and Environment, when related to Europe.

This section will describe the way this framework will be implemented in the 10 countries. What are the procedures that will take place and how the resulting data will be

processed and analysed. The projected deliverable is a report on the platformization of news in the 10 countries. That report will be produced by the Iscte-IUL team with the contribution of all the partners in the consortium. The step by step of framework implementation is represented on Figure 14.

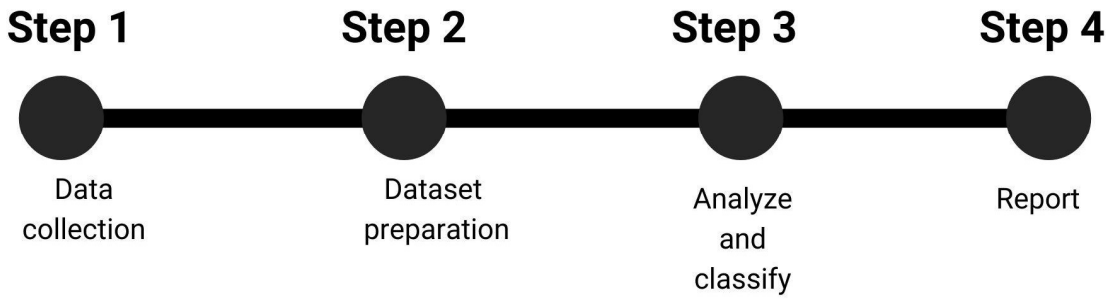


Figure 14: Step by step of framework implementation in the 10 countries.

4.1 Data collection

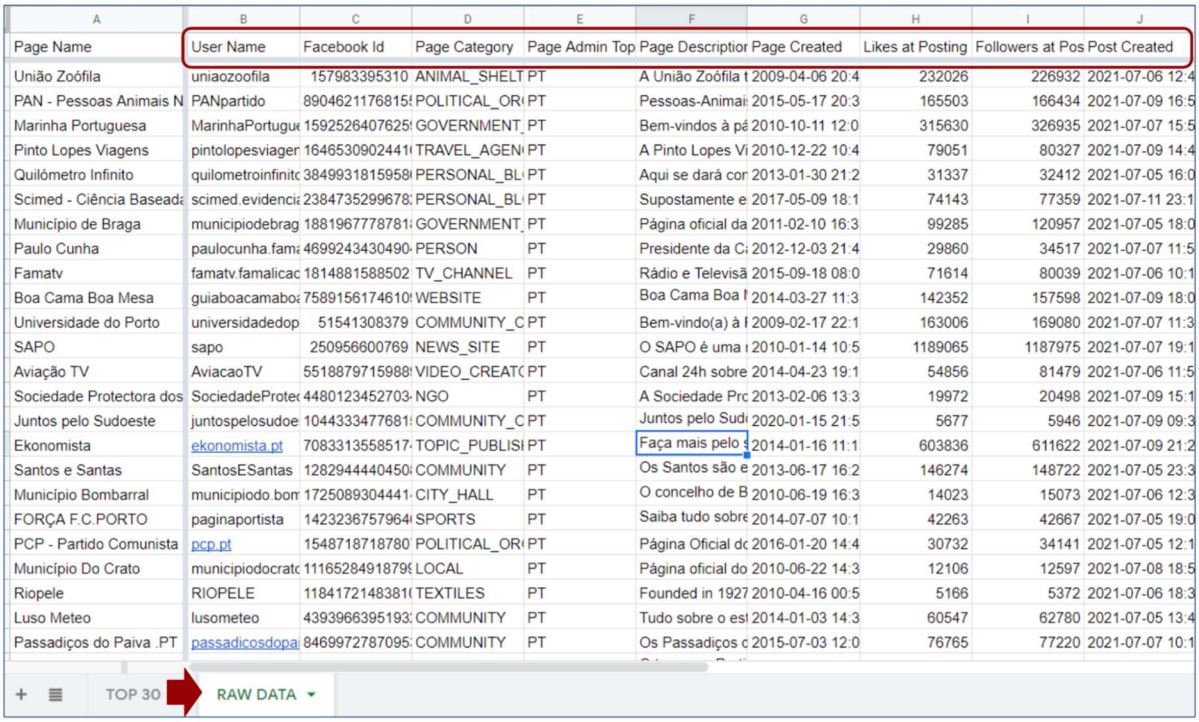
As detailed in the previous section, data collection will take place between September and November 2021. This will generate, for each country a total of 18 datasets. All datasets will have (as much as possible) similar structures and comparable data, both for dimension, country, platform and professional or non-professional content. The subsequent section describes how that data will be presented to the partners as well as what to do with it, in order to generate the final report about platformization in 10 countries that is the expected deliverable of WP2, task 2.2.

4.2 Dataset preparation

The team at Iscte-IUL will supervise and operationalize the extraction of data according to this framework. We will use the tools mentioned above and - through the authorized API's - extract the data for analysis. That data will then be uploaded for a shared server, accessible by all partners and organized in a tree-like structure: 1) country; 2) months; 3) dimensions; and 4) datasets.

Each dataset will have two different tabs: Raw Data and Top30. The **Raw Data** tab (see Figure 15) will exhibit the complete dataset as extracted from the corresponding social media platform, using the corresponding extraction tool. This dataset will have no edition other than the blocking of the first line and column. That means that each partner may work with these tabs, reordering the data by any of the criteria in the first line to observe different

perspectives on the data. This should be particularly relevant when considering the different metrics each extraction includes (again, those afforded by the API each tool uses). Each partner may also create news tabs to eventually perform more complex operations on the data for each dataset.



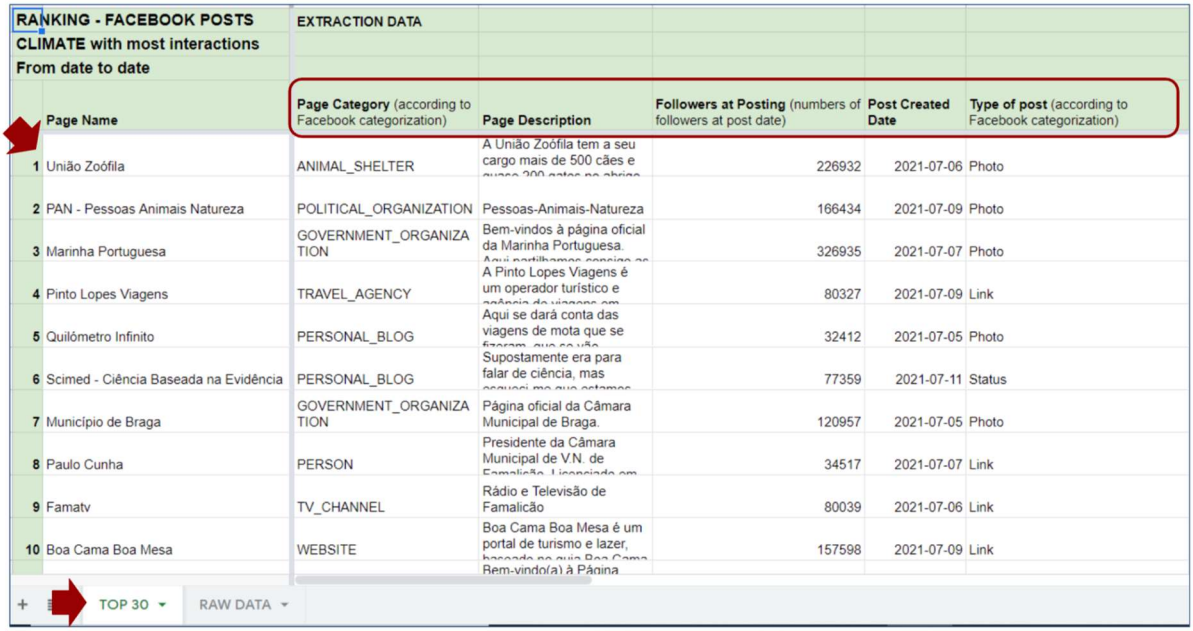
Page Name	User Name	Facebook Id	Page Category	Page Admin Top	Page Description	Page Created	Likes at Posting	Followers at Pos	Post Created
União Zoófila	uniaozoofila	157983395310	ANIMAL_SHELT	PT	A União Zoófila t	2009-04-06 20:4	232026	226932	2021-07-06 12:4
PAN - Pessoas Animais N	PANpartido	8904621176815	POLITICAL_OR	PT	Pessoas-Animai	2015-05-17 20:3	165503	166434	2021-07-09 16:5
Marinha Portuguesa	MarinhaPortugue	1592526407625	GOVERNMENT_	PT	Bem-vindos à pá	2010-10-11 12:0	315630	326935	2021-07-07 15:5
Pinto Lopes Viagens	pintolopesviager	1646530902441	TRAVEL_AGEN	PT	A Pinto Lopes Vi	2010-12-22 10:4	79051	80327	2021-07-09 14:4
Quilómetro Infinito	quilometroinfinito	3849931815958	PERSONAL_BL	PT	Aqui se dará cor	2013-01-30 21:2	31337	32412	2021-07-05 16:0
Scimed - Ciência Baseada	scimed.evidencia	2384735299678	PERSONAL_BL	PT	Supostamente e	2017-05-09 18:1	74143	77359	2021-07-11 23:1
Município de Braga	municipiodebrag	1881967778781	GOVERNMENT_	PT	Página oficial da	2011-02-10 16:3	99285	120957	2021-07-05 18:0
Paulo Cunha	paulocunha.fam	4699243430490	PERSON	PT	Presidente da C	2012-12-03 21:4	29860	34517	2021-07-07 11:5
Famatv	famatv.famalicac	1814881588502	TV_CHANNEL	PT	Rádio e Televisã	2015-09-18 08:0	71614	80039	2021-07-06 10:1
Boa Cama Boa Mesa	guiaboacamaboa	7589156174610	WEBSITE	PT	Boa Cama Boa M	2014-03-27 11:3	142352	157598	2021-07-09 18:0
Universidade do Porto	universidadedop	51541308379	COMMUNITY_C	PT	Bem-vindo(a) à U	2009-02-17 22:1	163006	169080	2021-07-07 11:3
SAPO	sapo	250956600769	NEWS_SITE	PT	O SAPO é uma p	2010-01-14 10:5	1189065	1187975	2021-07-07 19:1
Aviação TV	AviacaoTV	5518879715988	VIDEO_CREATO	PT	Canal 24h sobre	2014-04-23 19:1	54856	81479	2021-07-06 11:5
Sociedade Protectora dos	SociedadeProtec	4480123452703	NGO	PT	A Sociedade Pro	2013-02-06 13:3	19972	20498	2021-07-09 15:1
Juntos pelo Sudoeste	juntospelosudoe	1044333477681	COMMUNITY_C	PT	Juntos pelo Sud	2020-01-15 21:5	5677	5946	2021-07-09 09:3
Ekonomista	ekonomista.pt	7083313558517	TOPIC_PUBLISI	PT	Faça mais pelo s	2014-01-16 11:1	603836	611622	2021-07-09 21:2
Santos e Santas	SantosESantas	1282944440450	COMMUNITY	PT	Os Santos são e	2013-06-17 16:2	146274	148722	2021-07-05 23:3
Município Bombarral	municipiodo.bom	1725089304441	CITY_HALL	PT	O concelho de B	2010-06-19 16:3	14023	15073	2021-07-06 12:3
FORÇA F.C.PORTO	paginaportista	1423236757964	SPORTS	PT	Saiba tudo sobre	2014-07-07 10:1	42263	42667	2021-07-05 19:0
PCP - Partido Comunista	pcp.pt	1548718718780	POLITICAL_OR	PT	Página Oficial do	2016-01-20 14:4	30732	34141	2021-07-05 12:1
Município Do Crato	municipiodocrato	1116528491879	LOCAL	PT	Página oficial do	2010-06-22 14:3	12106	12597	2021-07-08 18:5
Riopele	RIOPELE	1184172148381	TEXTILES	PT	Founded in 1927	2010-04-16 00:5	5166	5372	2021-07-06 18:3
Luso Meteo	lusometeo	4393966395193	COMMUNITY	PT	Tudo sobre o est	2014-01-03 14:3	60547	62780	2021-07-05 13:4
Passadiços do Paiva .PT	passadicosdopa	8469972787095	COMMUNITY	PT	Os Passadiços d	2015-07-03 12:0	76765	77220	2021-07-07 10:1

TOP 30 RAW DATA

Figure 15 - Screenshot showing part of the complete data for each dataset extraction. There will be a dataset for each country/month/dimension/platform/user type.

The **Top 30** tab (see Figure 16) will comprise the 30 most relevant posts each month, ranked by the metric used for each platform (interactions in the case of Facebook; reach in Twitter and relevance in YouTube). It is on this tab that the first level of analysis and categorization, by partner in each country, will take place. This tab will detail the contents of each column and will display columns for the categorization of posts, according to a codebook that we detail in the next section.

At first, this Top 30 tab will display the 30 most important posts in each month. However, only up to 10 should be coded according to the codebook. The excess is purposeful and it has the goal of compensating for the eventual off-topic posts. When extracting posts or publications from social media using keywords it's possible that some of those posts will be off-topic. That will be the first goal of the codebook: to discard the posts that are off-topic. If less than 10 posts are considered on-topic, then news posts will be added, from the raw data, to complete the amount of up to 10 on-topic posts to categorize. This means that on-topic categorizable posts should be no less than 10 for each dataset.



	RANKING - FACEBOOK POSTS CLIMATE with most interactions From date to date	EXTRACTION DATA				
	Page Name	**Page Category** (according to Facebook categorization)	**Page Description**	**Followers at Posting** (numbers of followers at post date)	**Post Created Date**	**Type of post** (according to Facebook categorization)
1	União Zoófila	ANIMAL_SHELTER	A União Zoófila tem a seu cargo mais de 500 cães e	226932	2021-07-06	Photo
2	PAN - Pessoas Animais Natureza	POLITICAL_ORGANIZATION	Pessoas-Animais-Natureza	166434	2021-07-09	Photo
3	Marinha Portuguesa	GOVERNMENT_ORGANIZATION	Bem-vindos à página oficial da Marinha Portuguesa.	326935	2021-07-07	Photo
4	Pinto Lopes Viagens	TRAVEL_AGENCY	A Pinto Lopes Viagens é um operador turístico e	80327	2021-07-09	Link
5	Quilómetro Infinito	PERSONAL_BLOG	Aqui se dará conta das viagens de mota que se	32412	2021-07-05	Photo
6	Scimed - Ciência Baseada na Evidência	PERSONAL_BLOG	Supostamente era para falar de ciência, mas	77359	2021-07-11	Status
7	Município de Braga	GOVERNMENT_ORGANIZATION	Página oficial da Câmara Municipal de Braga.	120957	2021-07-05	Photo
8	Paulo Cunha	PERSON	Presidente da Câmara Municipal de V.N. de	34517	2021-07-07	Link
9	Famatv	TV_CHANNEL	Rádio e Televisão de Famalicão	80039	2021-07-06	Link
10	Boa Cama Boa Mesa	WEBSITE	Boa Cama Boa Mesa é um portal de turismo e lazer,	157598	2021-07-09	Link
			Bem-vindo(a) à Página			

TOP 30 RAW DATA

Figure 16 - Screenshot showing part of the Top 30 for each dataset extraction. Of these top 30 posts, 10 will be categorized according to the codebook.

4.3 Analyze and Classify

The goal of this framework is to analyse the platformization of news in the 10 countries that compose the EUMEPLAT consortium. In that regard, the framework will generate a sample of social media posts, both by professional and non-professional users of different social media platforms, about issues related to four different dimensions connected to Europe. That sample of posts is the corpus that will be analysed and categorized in each country and - by extension - in all 10 countries.

In detail, we will have datasets about Europe, Health, Economy and Environment (the last three as related to Europe), on three different social media platforms. On Facebook we will analyse and categorize: a) Professional news content on Facebook pages; b) Non-professional content on Facebook pages; c) Non-professional content on Facebook public groups. On Twitter we will analyse and categorize: a) Professional news content; and b) Non-professional content. And, on YouTube, we will analyze and categorize: a) Non-professional content.

The first step of analysis is to discard posts that are off-topic, as mentioned above. This means that all posts that will be finally considered for categorization are valid on-topic posts. The second step is to characterize all those posts according to a codebook that is currently being developed.

This approach means that we will be combining quantitative with qualitative methods. We are using quantitative criteria to extract the posts that compose our raw data and to select

those that will integrate our sample. Then, the categorization of the posts that compose that sample will be qualitative, according to a defined codebook.

The analysis enabled by that codebook starts with the content of each post but extends to several data points available as metadata to each post. This codebook will include categorias such as:

a) The format of the post: text, link, image, video, etc…;

b) The agent who posted it: institution, politician, TV host, anonymous citizen, influencer, etc…;

c) The subject matter: Institution, politician, TV host, anonymous citizen, influencer, etc…;

d) The Dimensions of Europeanization: cultural, economic, legal, etc…;

e) The sentiment towards Europe: negative, positive or neutral.

Same as similar queries will permit the comparison between results in different countries, the use of a similar codebook on all datasets is the methodological aspect that will allow comparison between the qualitative analysis of the sample of post in each country.

4.4 Report

For the report about the platformization of news in 10 countries - the deliverable expected for WP2, task 2.2 - each country will have freedom to explore the data, however within the limits imposed by data sample (determined by the queries and tools used and similar in all countries) and by the codebook (also similar in all countries).

Each country will be asked to produce a partial micro-report on their country of 5 to 10 pages, including possible tables, charts and images. The final report will include a supra-national analysis and methodology (redacted by the Iscte-IUL team,) with 10 to 20 pages (also including tables, charts and images), as well as the compilation of those national analyses.

The codebook is expected to be influential in what will compose the national and international reports, but the analysis in each country should address relevant questions such as these:

- Which of the 4 dimensions is more relevant in your country?
- What are the most frequent issues related to each dimension?

- What dimensions of Europe/Europeanization are most frequent?
- Who are the most prevalent agents in the discussion of these dimensions in your country?
- Professional or non-professional? In this case, what kind of agents?
- Which is the subject matter of the publications?
- What is the most used or popular format of the messages (video, image, audio, text…)?

With the results from these national as supra-national reports, we will be able to trace a picture of the state of the platformization of news in the 10 countries, as much as Europe is concerned. Plus, data will be able to be used to perform different types of analysis, combining correlations between different data points, able to respond to diverse research questions.

5 Data Management Plan – concerning WP2 – Fake News: Platformization of Journalism

The framework designed to study the platformization of news in 10 countries (WP2, task 2.1) and its implementation (WP2, task 2.2) will deal with data to a significant degree. Therefore, a Data Management Plan must be taken into consideration, both at the level of the collection of data as well as its subsequent storage and treatment. This section describes that Data Management Plan.

5.1 Description of the data

In the context of this framework for the study of the platformization of news in 10 countries, different types of data will be collected, from different sources (social media platforms) and using different extraction tools. Also, the data will be hosted in shared servers and will undergo some treatment, both by the extraction team and by the coding teams in each country.

5.1.1 Type of study

The main objectives of WP2 are related to the “transformations in journalism and news production”, and “concerns about undue political interference and fake news”. By means of a synoptic analysis of professional and user-generated news contents, expected to provide us with a better understanding of potentialities, limits and reliability of the process conducive to the platformization of journalism in Europe. To that end, we will survey both professional and

non-professional news production on social media. The survey will be conducted in the ten countries represented in EUMEPLAT network.

WP2 will be based on a synchronic investigation of contemporary information systems. One of the main changes affecting the media ecosystem has to do with recent evolution in news production and consumption, usually referred to as the "platformization of news". With this respect, platformization of news triggered a real revolution, leading to both positive and negative externalities – on the one hand, civic participation and the widening of the communication arena; on the other hand, the spread of fake news and the so-called polarization effect.

WP2 will analyse the evolution of news production, circulation and distribution in digital environments, and in the ten countries represented in the consortium.

One of the WP main results will be an analysis of anti-European misinformation on social media platforms (Facebook, Twitter, YouTube) and provide data-driven policy recommendations and countermeasures.

5.1.2 Type, nature and consistency of data

The WP will collect data from public entities such as Facebook pages/groups, public Twitter accounts and YouTube public channels. To achieve that goal, the three major social media platforms – Facebook, Twitter and Youtube – will be involved for collecting the personal data necessary for this Work Package. As indicated in the previous paragraphs, as far as perosnal data may be collected and processed, such data will be managed according to the GDPR. All data are collected from public domains and accounts. Only personal information that users of social media platforms choose to make publicly available may be collected.

Regarding the categories of personal data processed. We specify the following:

No sensitive data will be collected insofar as that type of data is not included in the public data provided by the APIs.

No genetic data are or will be collected for the research.

No biometric data are or will be collected for the research.

No data concerning health are or will be collected for the research.

No children are or will be directly involved in the research. For instance, we precise that the tool CrowdTangle has mechanisms in place not to collect or store information from children under the age of 13. It also has mechanisms for detecting and erasing information that may be inadvertently collected from children under 18 years of age. In case children are indirectly involved, the sources of data being Social Media platforms, the EUMEPLAT researchers will not use contents produced or uploaded by children before the so called "digital age of consent"

[GDPR art. 8], which is the minimum age a person must be for social media companies to collect their data. According to the same article, Member States are allowed to lower the age limit to 13 years, which is the case in Czech Republic, Spain, and Sweden. Therefore, the research could involve public posts of children between 13 up to 18 years old: withint Task 2.1 their posts will be stored in a secure database in the restricted area of the EUMEPLAT website, located in Italian servers and accessible only by researchers provided with a specific username and password.

The format of the data will be tables (CSV).

5.2 Data collection and generation

The research is directed at the production, distribution and consumption of news and related content in 10 countries individuated by the Consortium. Collection of data by he means indicated below will be carried out during a limited period of time, by ISCTE-IUL. The collection phase will last three months. The collection will occur via the public and authorized APIs (Application Programming Interfaces) made available by the platforms, as detailed below. No data will be scraped outside those authorized APIs. Connection to those APIs will be made through specific and reliable tools: all data from Facebook will be collected through CrowdTangle; all data from Twitter will be collected using Brandwatch; and all data from YouTube will be collected with YouTube Data Tools.

Since the collected data will not directly obtained from the data subjects/social media platforms users, the most appropriate way to provide information according to Art. 14 GDPR will be defined. At this stage, providing a dedicated privacy policy on the project's website seems the most adequate option. In addition, according to Art. 26, par. 2, GDPR, the joint controllership agreement will be made available to the data subjects, through the project's website. Sharing this information will be important also for compliance with the privacy by desing principle, as set out in Art. 25 GDPR.

5.2.1 Facebook

The entire data collection process on Facebook is performed exclusively through CrowdTangle, a public insights tool owned by Facebook that operates via the available public Graph API[9]. CrowdTangle uses only publicly available data and exclusively tracks public content. Data is downloaded from Facebook pages/groups that are public entities. We abide

[9] https://developers.facebook.com/docs/graph-api/ and https://developers.facebook.com/docs/graph-api/reference/v2.10/comment

by the terms, conditions, and privacy policies of Facebook. We have no access to information about the users who reacted/commented to Facebook content on public pages/groups. For each public post, we have the numeric ID and the name associated to the publishing account, the message contained within the post, the date and time in which the post was initially published, the type of post (link, photo, video etc.), the link attached to the post, the post ID, the "story" description associated to the post, the aggregated number of reactions, comments and shares, and the numeric ID associated to the page in which the post is published. Moreover, users' *reactions* include the number of reactions each post got ("angry", "hah", "like", "sad", "wow"). We abide by the terms, conditions, and privacy policies of Facebook

5.2.2 Twitter

Within Task 2.2, the entire data collection process on Twitter is performed using a Brandwatch account (owned and operated by Iscte-Iul). Brandwatch[10] is a commercial information retrieving company that operates exclusively within the framework of the Twitter API[11], which is publicly available. With respect to Brandwatch Master Subscriptions Agreement, and namely to Article 5.3, we will own intellectual rights on retrieved information, while no commercial use of those data is allowed [article 3.2]. We use only publicly available data. Users with privacy restrictions are not included in our dataset. Data is downloaded from Twitter accounts that are public entities.

Data will include public tweets made on the timelines of public users that correspond to a search query, as well as the number of retweets, replies and mentions corresponding to those tweets. No personal data from the users will be downloaded other than that which is publicly available through the API. We abide by the terms, conditions, and privacy policies of Twitter.

5.2.3 YouTube

The entire data collection process on YouTube is performed using the YouTube Data Tools developed by DMI (Digital Methods Initiative) at the University of Amsterdam, which are publicly available and operate exclusively by means of the YouTube Data V3 API[12], which is also publicly available. We used only publicly available data. Users with privacy restrictions are not included in our dataset. Data is downloaded from YouTube channels that are public entities. We abide by the terms, conditions, and privacy policies of YouTube.

---

[10] https://www.brandwatch.com/blog/brandwatch-and-the-gdpr-what-you-need-to-know/

[11] https://developer.twitter.com/en/docs/api-reference-index

[12] https://developers.google.com/youtube/v3/getting-started

Data will include all data relative to the published public videos made on public channels corresponding to a search query, as well as the number of views, likes, dislikes, favorites and comments corresponding to those videos. This would include the timestamp of the video, it's title and caption and tags. No personal data from the users will be downloaded other than that which is publicly available through the API.

5.3 Data Processing

5.3.1 Extraction

Data will be extracted using the tools referred above, through the publicly available APIs and respecting the terms, conditions, and privacy policies of each online platform.

After the clearance provided by the Ethical Committee, data collecting will start, for a three-month period. Data will be collected in the ten countries represented in the Consortium, and namely: Italy [Italian], Germany [German], Greece [Greek], Belgium [French and German], Portugal [Portuguese], Sweden [Swedish], Turkey [Turkish], Spain [Spanish and Catalan], Czech Republic [Czech], Bulgaria [Bulgarian]. With the exception of UNIMED, each partner will receive data and analyze them, with Bilkent University of Ankara putting the case of data transfer outside the European Union.

Since we are extracting data only from Facebook pages/groups that are public entities, Twitter accounts that are public entities and YouTube public videos made on public channels, it is reasonable to think that all data will come from institutions, organizations or people that made it public. We will not collect comments or any other form of personal inputs into pages or groups (posts in groups are not identifiable). Beyond the public nature of the content itself, what is more relevant, both social media terms of use and Art.89 GDPR allow the use of data for scientific research purposes. No indirect or commercial reuse of those data is allowed.

5.3.2 Analysis

The data analysis techniques which will be adopted include: content analysis, frequency analysis, and audience engagement analysis. The results of the analysis will be presented in aggregation and no personal information will be disclosed.

5.4 Managing and storing data

5.4.1 Data storage and transfer

Data and results will be stored in the restricted area of the project website and made available solely to the authorized partner researchers. All data will be stored in servers located in Italy.

Data destination is EU27 plus Turkey. As a consequence, access to data base may imply a transfer of personal data outside the European Economic Area (EEA), with the application of artt. 45 and ss. GDPR. As a consequence, access to Bilkent University will be authorized upon identification of appropriate safeguards according to artt, 46 and ss. GDPR. At this stage, subscription of Standard Contractual Clauses seems the most appropriate choice.

Access to data on the storage drive is subject to authentication using username and password. Strong authentication creditials are required according to Art. 32 GDPR. Only researchers involved in the project and duly appointed according to art. 29 GDPR will have access to the data thanks to a separated and restricted area, accessible only upon authentication. Data will be stored on secure servers (as indicated) and only institutional computers and acoounts will be used for the Project.

Therefore, to ensure appropriate protection to the personal data collected, there will not be authorized any storage on personal laptop computers. The best practice for secure storage of data will be respected, such as the recourse to encryption.

In order to comply with both GDPR rules and Open Science principles, data will not be publicly shared, while they will be available for researchers upon request. The Steering Committee will be in charge for the evaluation of the access request, for reasons including: data checking for peer-review evaluation of research outputs; data comparison for similar research activities in the field of media studies and European studies.

5.4.2 Data management and storage facilities

Data destination is EU27 plus Turkey. Data will be stored in a shared cloud drive available to the partners and created specifically for this purpose.

Access to data on the storage drive is subject to authentication using username and password. Only researchers involved in the project will have access to the data. This cloud drive will be administered by Iscte-IUL team following the best practices and standards available.

5.4.3 Data preservation strategy and standards

In compliance with the GDPR EU 679/2016 law, data will be shared only among the participants to the project; they are therefore closed access.

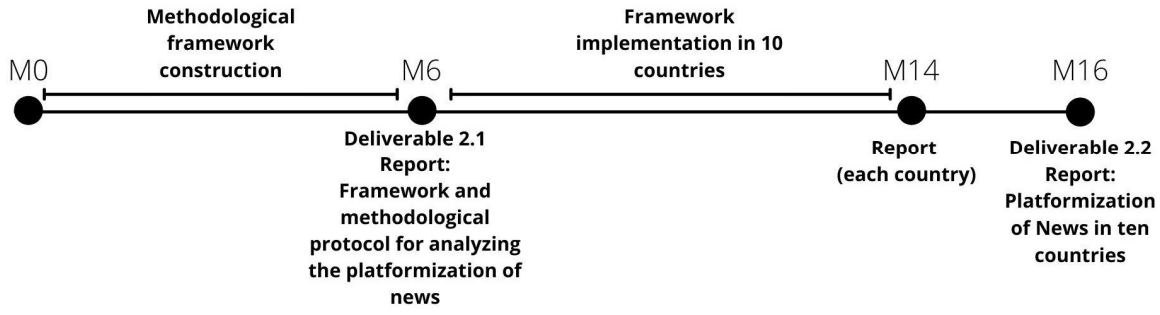
5.4.4 Main risks to data security

No significant risk is expected. Backup copies of data will be done according to the best security procedures and practices.

5.5 Responsibilities

Data collection, processing, management and storage are carried out by Iscte-Iul. Data is collected, processed and managed under the responsibility of the project principal investigator Dr. Cláudia Álvares.

6 Timetable



References

Brandwatch for Research. (2021). Retrieved 21 September 2021, from https://www.brandwatch.com/p/brandwatch-for-research/

Carpentier, N. (2021). The European Assemblage: A Discursive-Material Analysis of European Identity, Europaneity and Europeanisation. Filosofija. Sociologija, 32(3). doi: 10.6001/fil-soc.v32i3.4495

Coromina, O. & Molina, A. P. (2018). Reconstructing memory narratives on Facebook with Digital Methods. Culture & History Digital Journal 7(2), e014 eISSN 2253-797Xdoi: https://doi.org/10.3989/chdj.2018.014

CrowdTangle. (2021). What data is CrowdTangle tracking?. Retrieved 21 September 2021, from https://help.crowdtangle.com/en/articles/1140930-what-data-is-crowdtangle-tracking

Digital 2021. (2021). Retrieved 20 September 2021, from https://wearesocial.com/digital-2021

Eurobarometer (2020). Standard Eurobarometer 93 - Summer. Retrieved from

https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Survey/getSurveyDetail/instruments/STANDARD/surveyKy/2262

Glaser, B., & Strauss, A. (1967). The Discovery of Grounded Theory Strategies for Qualitative Research. Mill Valley, CA Sociology Press.

Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C. T., & Nielsen, R. K. (2021). Reuters Institute Digital News Report 2021. Reuters Institute for the Study of Journalism.

Omena, J.J. (2019). Métodos Digitais Teoria - Prática - Crítica. Lisboa: ICNOVA – Instituto de Comunicação da Nova, p. 5-15 10 p. (Livros ICNOVA).

Poell, T., Nieborg, David & Van Dijck, J. (2019). Platformisation. Policy review. Volume 8. 10.14763/2019.4.1425.

Rieder, B. (2015). YouTube Data Tools (Version 1.22) [Software]. Available from https://tools.digitalmethods.net/netvizz/youtube/.

Rogers, R. (2017). Foundations of Digital Methods: Query Design. In: Mirko Tobias Schäfer, Karin van Es (Hg.):The Datafied Society: Studying Culture through Data. Amsterdam: Amsterdam University Press 2017, S. 75–94. DOI: https://doi.org/10.25969/mediarep/12536.

Salmons, J. (2016). *Doing qualitative research online*. Sage.

Statcounter. (2021). Social Media Stats Europe | Statcounter Global Stats. Retrieved 30 September 2021, from https://gs.statcounter.com/social-media-stats/all/europe

Van Dijck, J., Poell, T., & De Waal, M. (2018). The Platform Society. Oxford: Oxford University Press USA - OSO.

Williams, Raymond. 1975. Keywords: A Vocabulary of Culture and Society. London: Fontana.

Stevenson, A. & Lindberg. C (Ed) (2015). New Oxford American Dictionary (3 ed.).Oxford University Press.

ANNEXES

Annex 1 - Examples of keywords exploration and query construction.

Example - Search for initial keywords related with Covid in the last 12 months:

Keywords relacionadas com Covid nos útimos 12 meses			
FACEBOOK (1)	**Twitter**	**Twitter Hastags**	**Google Trends**
Covid-19	Covid-19	#covid19pt	covid
pandemia	casos	#coronavirus	covid 19
Portugal	Portugal	#portugal	portugal covid
casos	#covid19	#noticia	covid 19 portugal
Saúde	pandemia	#cm	covid hoje
pandemia de covid-19	mortes	#covid	covid-19
pessoas	saúde	#saúde	covid casos
país	pessoas	#pandemia	covid hoje portugal
mundo	mortos	#observador	covid dgs
infetados	infetados	#coronavírus	covid 19 hoje
mortes	vacina	#algarve	covid sintomas
últimas	Governo	#estamoson	teste covid
Lisboa	mundo	#atualidade	covid em portugal
Governo	regista	#covid—19	covid hoje casos
estado	últimas	#alentejo	casos covid portugal
presidente	milhões	#sns	covid 19 hoje portugal
últimas 24	país	#mundoaominuto	casos de covid
mortos	casos de Covid-19	#dgs	covid19
milhões		#saude	vacina covid
medidas			covid-19 portugal
crise			noticias covid
dias			covid 19 dgs
países			sintomas covid 19
combate			medidas covid
dados			covid numeros

Example - Query on Health issues, as related to Europe, in portuguese:

```
(       "acidente vascular cerebral"          OR
        "administração regional de saúde"  OR
        "agência europeia do medicamento"          OR
        "certificado digital"          OR
        "cuidados intensivos"          OR
        "direção geral da saúde"   OR
        "direção geral de saúde"   OR
```

"direção-geral da saúde" OR

"direção-geral de saúde" OR

"doentes de risco" OR

"estado de calamidade" OR

"estado de emergência" OR

"estirpe indiana" OR

"isolamento profilático" OR

"população de risco" OR

"teste rápido" OR

"testes rápidos" OR

"tratamento médico" OR

#coronavirus OR

#coronavírus OR

#covid19 OR

#DGS OR

#estamoson OR

#fiqueemcasa OR

#pandemia OR

#saude OR

#saúde OR

#sejaumagentedesaúdepública OR

#SNS OR

#umconselhodaDGS OR

#VacinaçãoCovid19 OR

ambulância OR

ambulatório OR

ansiedade OR

anticorpos OR

ARS OR

astrazeneca OR

casos OR

clínica OR

clínicas OR

clínico OR

clínicos OR

co-morbilidades OR

confinamento OR

confinamento OR

coronavirus OR

coronavírus OR

covid OR

covid-19 OR

covid19 OR

desconfinamento OR

DGS OR

doença OR

doenças OR

doente OR

doentes OR

enfermaria OR

enfermeira OR

enfermeiras OR

enfermeiro OR

enfermeiros OR

epidemia OR

epidémica OR

epidémico OR

EstamosOn OR

farmacêutica OR

farmacêutico OR

farmácia OR

farmácias OR

hospitais OR

hospital OR

hospitalar OR

imunidade OR

imunização OR

incidência OR

infetada OR

infetadas OR

infetado OR

infetados OR

inoculação OR

máscara OR

máscaras OR

médica OR

medicamento OR

medicamento OR

médicas OR

medicina OR

medicina OR

medicinal OR

médico OR

médicos OR

MySNS OR

OMS OR

paciente OR

pacientes OR

pandemia OR

pandémica OR

pânico OR

Pfizer OR

quarentena OR

saudáveis OR

saudável OR

saúde OR

SNS OR

terapêutica OR

terapia OR

trombose OR

vacina OR

vacinada OR

vacinadas OR

vacinado OR

vacinados OR

virologia OR

virologista OR

vírus)

AND

(europa OR

europeu OR

europeia OR

europeus OR

europeias OR

UE)

“AND” and “OR” operators

The OR operator broadens the ensemble of media objects we can collect (posts, tweets, etc) whereas AND limits the ensemble of media objects we can collect. The OR operator should be used to gather social media objects that span across the overall range of an issue (e.g. Health); the AND operator, on the other hand, should be used when we want to focus on a specific issue (e.g. vaccines AND europe).

	Pros	Cons
OR operator	> Broaden the spectrum of issues (or issues within an issue) gathered > Results closer to “what people are talking about” (“get the pulse” on social media) > Easier detection of relevant and (sometimes) not obvious influencing actors > Less influence from subjectivity (in determining keywords, filters and inclusion criteria)	> Too broad spectrum of issues (or issues within an issue) > Results may me more distanced from theoretical categories (making it difficult to interpret theoretically) > Possibility of very prominent media objects or social media actors “dwarfing” smaller niche but relevant posts or actors > Greater presence of “off-topic” content
AND operator	> Filter the results to a specific issue (more “to-the-point” content) > Less “noise” (false positives regarding the issue we want to adress) > Less (or null) “off-topic” content > More suitable to identify relevant actors regarding a specific issue > Easier to connect to theoretical concepts	> Construct an image of social media discourse that is more distant from social media “actual” discourse about an issue > Not having the actual “pulse” on the issue > With our current software for extraction of data from Facebook (Crowdtangle), AND operator does not allow a monthly ranking of pages/groups with most posts/most interactions on the issue.

Annex 2 - WP2 tasks 2.1 and 2.2

Task	Framework and general question	Specific research question	Research tasks	Possible indicators [a proposal]	Measurable results according to the proposal
2.1 A methodological framework for analyzing the platformization of news [M 1-6] Leader: ISCTE	[Methodology]	[Methodology]	[Setup methodological instruments; Test methodological instruments]	[Setup of a methodological protocol]	[Deliverable 2.1 Report: Framework and Methodological Protocol]
2.2 Platformization of news in ten countries [M 6-16] Leader: ISCTE	Which are the most relevant issues in European media, and how are citizens debating about them?	Which debate is taking shape at the intersection of bottom-up [professional] and top-down [user-generated] communication in social media platforms, in the ten countries?	Analysis of social media posts from relevant media and citizens in the ten countries [each partner in its own country, at least]	Analysis of professional social media posts related to the selected issues, in each country; Analysis of amateur social media posts related to the selected issues, in each country;	At least 500 selected cases of news Production analyzed [Deliverable 2.2 Report: Citizen Journalism in Ten Countries

Get in touch

info@eumeplat.eu

www.eumeplat.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004488