free litter aT

ADVANCING TOWARDS LITTER - FREE ATLANTIC COASTAL COMMUNITIES BY PREVENTING AND REDUCING MACRO AND MICRO LITTER

Retrieving information on beach clean-up activities from social network sites. Socio-economic impact.

WORK PACKAGE 1. ACTIVITY 4, TASK 2

Interreg Atlantic Area
Co-funded by the European Union
free litter aT

Work package	WP 1. PREVENTION BY IMPROVING WASTE MANAGEMENT AND RECYCLING
Activity and task	ACTIVITY 1, TASK 2. INVESTIGATION OF ENVIRONMENTALLY RESPONSIBLE BEHAVIOUR ON SOCIAL NETWORKS BY USING DATA OBTAINED FROM THE SOCIAL NETWORK X
Date	23 / 09 / 2024
Version	3
Author/s:	OTERO, P., CACABELOS, E., BORGES, J. AND GAGO, J.
Participants	REVISED BY SEPÚLVEDA, P., CLOUX, S., PESTANA, N., JONES, O.

Table of contents

1. INTRODUCTION

Modern society cannot be understood without considering the role of the internet, where an enormous amount of information is shared. Social networks have become an essential part of this digital ecosystem, through which millions of citizens share text, images and videos in real time. This massive amount of digital information also includes citizens' outdoor activities, which have opened up new opportunities to study human interactions with nature (Arts et al., 2015; Di Minin et al., 2015; Hausmann et al., 2018) at an unprecedented spatial and temporal scale (Ladle et al., 2016). Online publications also allows us monitor environmental concerns (Cooper et al., 2019; Correia et al., 2017; Otero et al., 2025, 2024, 2021; Roll et al., 2016) and better understand ecosystem services related to leisure and culture (Gliozzo et al., 2016; Richards and Friess, 2015; Richards and Tunçer, 2018), among others. Importantly, this information can be accessed transparently, after volunteers share relevant content on social media.

Extracting valuable information from internet content poses challenges due to the huge amount of information and to the use of informal language. To address these challenges, it's crucial to develop methods that can effectively retrieve data from both text and accompanying multimedia. Additionally, obtaining location information without compromising user privacy is essential. The rapid growth and ephemeral nature of social media posts necessitate the implementation of automated systems for information capture (Toivonen et al., 2019) and subsequent classification to streamline human validation. Achieving this requires the creation of algorithms that integrate natural language processing with image analysis techniques.

The social media platform formerly known as Twitter (2007–2023), now called X Corp (2023–present), is well-known for its emphasis on current events and related discussions, often with a light-hearted approach. Users on this network share text, multimedia, and geo-located content, making the platform a rich source of information. In contrast to other popular social media platforms such as Facebook, Instagram, or TikTok, X Corp's comprehensive and well-documented API enables the collection of more extensive data, making it particularly valuable for various studies, including real-time global monitoring. This feature allows for more diverse datasets and enhanced potential for identifying worldwide patterns and trends. Consequently, it has been extensively utilized in numerous fields, including social studies (Budenz et al., 2020; Makita et al., 2021; Milani et al., 2020; Plackett et al., 2020; Teoh et al., 2018; Vraga et al., 2018; Yoosefi Nejad et al., 2019), environmental research (Otero et al., 2024, 2021; Ruiz-Frau et al., 2020), health investigations (Dang et al., 2018; Shimkhada et al., 2021), and emergency alert analyses (Barker and Macleod, 2019; Martínez-Rojas et al., 2018; Zhou et al., 2021), among others.

It is worth noting that many posts on X Corp may originate from other social networks and vice versa, such as Instagram (Heikinheimo et al., 2018), resulting in the same

content being distributed across multiple social media platforms. Instagram, owned by Meta Platforms, is a photo and video sharing social networking service. Users can upload content, apply filters for editing, and categorize it using hashtags and location tags. The platform enables users to explore content from others by searching tags and locations, view popular posts, appreciate photos with likes, and create a personalized feed by following accounts of interest. It is important to note that the search platform tool works differently in both web and mobile apps, offering more options through the mobile app. Moreover, the list of results is personalized for each individual user by Instagram's algorithm.

This report presents the design and implementation of an algorithm to obtain X and Instagram data related to beach and coastal clean-up activities. The objective is to obtain a dataset with information on this type of activities in the 4 countries of the Atlantic Area programme. This dataset will provide valuable information for understanding environmental awareness of the marine litter problem, as well as quantifying the social and economic impact associated with the development of beach and coastal clean-up campaigns.

2. METHODOLOGY

2.1. Algorithm for X data mining

An algorithm was designed in Python programming language to acquire tweets through the Twitter API v2 (https://developer.twitter.com/en/docs/twitter-api). To access the largest possible dataset, the 'Basic' paid access has been used, which allows to create 2 environments and to retrieve up to 10K tweets per month, along with a number of additional features that make it more suitable for research purposes (terms and conditions in force in October 2024).

The initial query is based on a set of keywords and hashtags, which are listed in Table 1. Tweets announcing the call for beach clean-up activities are usually accompanied by multimedia content, so images or videos are required in the tweet. The query returns exclusively tweets posted in the last 7 days.

TABLE 1: Keywords and hashtags used to retrieve data	
FRENCH	(NETTOYAGE or #NETTOYAGE or "RAMASSAGE DE DÉCHETS" or "RAMASSAGE DES DÉCHETS" or "COLLECTE DE DÉCHET" or "COLLECTES DE DÉCHETS" or "COLLECTES DES DÉCHETS") + (CÔTES or PLAGE or PLAGES or BALNÉAIRE or #CÔTES or #PLAGE or #PLAGES)
ENGLISH	(CLEAN-UP or CLEAN-UP or CLEAN UP) + (BEACH or BEACHES or COAST or HARBOUR or STRAND or BAY or COVE)
PORTUGUESE	(LIMPEZA or #LIMPEZA) + (COSTA or #COSTA or PRAIA or #PRAIA or PRAIAS OR #PRAIAS)
SPANISH	(LIMPIEZA or #LIMPIEZA) + (PLAYA or PLAYAS or CALA or COSTA or LITORAL)

Above keywords and hashtags were selected after an initial supervision of related tweets. The query to the X API includes, in addition to these words, other filters that allow to narrow the search and that are not shown in the table.

As is typical of big data, information extracted from this social network is characterized by its vast volume and significant diversity. This necessitates processes for systematic cleaning and filtering to concentrate on pertinent information (Poorthuis and Zook, 2017). Thus, the initial query to the API also looks for undesirable emoticons, or emoticons with double meanings, which are used as filters to narrow down the search. As the number of characters in the query to the API is limited, this first filter is not as extensive as would be desirable. Therefore, following filters are applied directly to the downloaded set of tweets.

Among the information accessible, the downloaded dataset contains not only the text of the tweet, but also the associated multimedia content and metadata. Available metadata may encompass the location (when available) and time of the tweet's publication, certain data on the user's profile, the unique identifier of the conversation thread held, and number of retweets and likes, among other data. Metadata also includes the position from which the tweet was posted in case the user had the location of the device active at the time of publication. Some tweets also have aggregated information about the estimated position of the user (city, state, country). All this information is employed to further refine the dataset, discarding tweets that do not meet the defined criteria:

- Tweets containing media in gif format, usually related to *memes* (https://en.wikipedia.org/wiki/Internet_meme).
- Tweets in response to other users or retweets.
- Tweets with similar content (same text) than others in the dataset.

- Tweets containing words that have been identified with a high likelihood of not being targeted tweets, for example, ‘Costa Rica’, ‘Mexico’, ‘Argentina’, ‘sex’, ‘nudist’, ‘adopts’, ‘cats’ and so on (note that we have shown the translated words to English, not the original words used in the algorithm). Therefore, the algorithm checks the text of the tweet against a prebuilt bag of words.

The algorithm also tries to obtain the location of the tweet from the tweet's metadata. However, the approach taken differs depending on the language and the country from which we want to obtain the information.

For those tweets retrieved with the Spanish query, it checks if any beach on the Spanish coast is mentioned in the text. To do so, it performs a search for the text ‘playa’ (beach) or ‘cala’ (cove) and, if available, analyses the set of successive words.

The goal is to find a valid beach name and check its existence using the ESRI API (https://opendata.esri.es/datasets/playas-españolas/api). If the location is not successful in the ESRI catalogue, then it tries the search in OpenStreetMap (OSM), using the Nominatim 4.2.3 service for this purpose. Nominatim always tries to return a collection of the best matches and does not necessarily have to match a coastal location. The algorithm also checks that the type of location corresponds to one of the following natural types: beach, cape, strait, coastline, reef or bay (https://wiki.openstreetmap.org/wiki/ES:Key:natural?uselang=es). This additional information is also provided by the OSM service. It should be noted that while the ESRI catalogue offers exclusively Spanish beaches (for this reason, the ESRI API is used first), OSM returns worldwide locations. Sometimes this entails an added complexity since there are beaches with the same name in Spain and in other countries, in particular in Latin America. For this reason, it is necessary to exclude all those locations that do not correspond to the Spanish coast, both peninsular and insular. This process is carried out by defining a bounding box around our study area. For those beaches within the national territory that share the same name, the location can be inferred from the proximity to the location declared in the user profile (if present).

To optimize filtering and geolocation with the Portuguese query, a dictionary of beaches was built based on the official tourism website of Portugal (https://www.visitportugal.com) and dedicated websites (e.g., https://www.praiasdeportugal.com). Subsequently, beaches were geolocated with OSM service and stored in a separate dictionary. Checking the dictionary is computationally less expensive than recurrently querying to OSM service. Beaches in Brazil with no correspondence in Portugal are used to improve the filtering process.

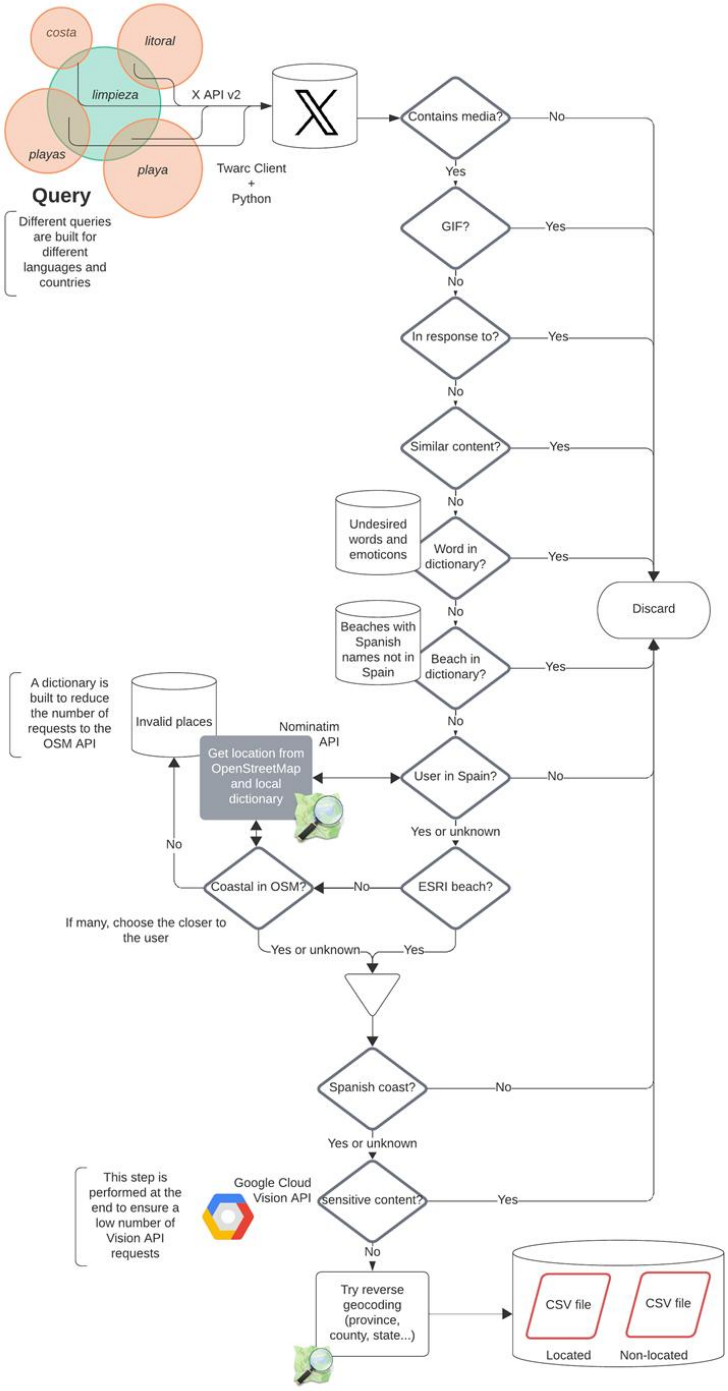

FIGURE 1: Architecture diagram of the X data mining for those tweets in Spain and written in Spanish. Similar although not equal implementations are done to retrieve tweets in France, Ireland and Portugal.

The strategy described above was also done with the French query. In this case, the list of beaches was obtained from a specialized website focused on French beaches, offering a comprehensive guide to both popular and hidden beaches across mainland France and its coastal regions (https://www.plages.tv) and the official tourism website of France, which provides detailed information about tourist spots, including beaches from different regions such as Brittany, Normandy, Aquitaine, and others (https://www.france.fr/). As part of the filtering process, tweets with direct reference to other French-speaking countries are discarded.

Similarly, the geolocation of Irish beaches and coastal locations is performed against a predefined dictionary. This dictionary was based on names retrieved from the official tourism website for domestic holidays in Ireland (https://www.discoverireland.ie/) and the list of beaches in the water quality program of the Environmental Protection Agency (EPA), and available at https://www.beaches.ie/.

Due to the existence of inappropriate images in some tweets, a filtering is performed using the Google Cloud Vision API SafeSearch tool (https://cloud.google.com/vision). In this case, content classified as adult or spoof, with a 'likely' or 'very likely' likelihood is discarded; this bucketized representation of likelihood, intends to give clients highly stable results across model upgrades. This filtering is done near the end of the process and not before, in order to reduce the number of API connections.

Finally, a virtual server was configured in order to execute, on a daily basis, the Python script containing the algorithm. As a result, a database of 'potential' tweets is obtained with a strong likelihood of referring to beach clean-up activities on the coasts of Spain, Portugal, Ireland and France. Dataset files are stored in both the server and in the cloud, with access by partners of the project upon request.

The developed algorithm and the list of beaches in the Atlantic Area countries and used in the algorithm are publicly available at: https://github.com/PabloOtero/freelitterat

To test the validity of the algorithm and its application in the Atlantic Area, data were retrieved during 3 months, from July 15th to October 15th, 2024 (see Results section). Datasets were manually supervised to identify real citizen beach clean-up initiatives. Table A in the Supplementary material shows some examples of the supervision process. Here, clean-ups carried out by city councils, public bodies, or specialized companies are excluded. Awareness-raising announcements that do not directly involve beach clean-up activities are also excluded.

2.2. Instagram data mining

The Instagram API allows users to obtain and publish media content, manage and respond to comments referring to that content, identify media content where other Instagram users have mentioned them, search for media content with hashtags, and obtain basic metrics and metadata about other Instagram creators and businesses. On November 2024 (during the design of the study), the access presented the following limitations:

- Request allows a maximum of 30 unique hashtags on behalf of a business account or Instagram creator within 7 days. When you query a hashtag, it will be deducted from this limit for 7 days. Subsequent queries regarding the same hashtag within this timeframe will not be deducted from your limit; nor will they reset the initial 7-day query timer.
- Request for hashtags in stories (not 'media' or publications) are not supported.

In order not to exceed the limit of hashtags, and as a first step, a search was made through Instagram's web interface to identify which hashtags were most recently used in reference to beach clean-up activities (Table 2). Hashtags written in any of the official languages of the Atlantic Area countries were searched.

However, the algorithm of the web search interface did not always provide consistent results for the same request. Sometimes it provided an exact number of results and sometimes only an upper limit (e.g. 'less than 100 results'). Results also varied when the query was performed on different dates.

It is also unclear to what date range the total results correspond or whether they are customized by the algorithm based on the user's profile who is performing the search. Likewise, it does not distinguish between uppercase and lowercase in characters when searching for hashtags, but it does respond differently with the presence of accents.

Posts containing a hashtag from Table 2 were collected as long as they met two requirements: i) they were most directly related to beach clean-up (no asterisk) and ii) the web search tool for recent posts returned a total of more than 200 results. Although data about post time, user nickname, associated image and geolocation when available, were initially stored, the final dataset was adequately anonymized.

TABLE 2: Hashtags commonly used in Instagram related to clean-up activities and their counts as provided by the Instagram web search interface on 5th November 2024

SPAIN (includes Spanish, Galician and Catalonian)	
#limpiezadeplayas	25482
#playalimpia	13958
#limpiezadeplaya	10968
#basuramarina	6161*
#limpiezaplayas	4189
#limpezadepraias	2041
#lixomariño	644*
#limpezapraia	626
#limpiezacostera	624
#netejadeplatges	233
#limpezapraias	203
#netejaplatja	148
#netejadeplatja	142
#limpiezacostas	88
#limpiezalitoral	63
#netejaplatges	62
#limpiezacosta	28
#limpezasimultáneadepraias	19
#limpezasimultaneadepraias	18
#limpiezasimultáneadeplayas	16
PORTUGUESE	
#limpezadepraias	2041
#limpezapraia	626
#limpezapraias	203
ENGLISH	
#beachclean-up	616536
#beachclean	226416
#clean-upday	59767*

#marinelitter	46971*
#coastalclean-up	+42000
#cleanthebeach	+23300
#EUBeachCleanup	+5000
#beachclean-upday	+1000
#coastclean-up	539
#beach_clean	264
#coastclean	241
#beach_clean_up	232
#beach_cleaning	190
#beach_clean-up	180
#coastcleaning	104
#beachclean-upchallenge	+100
#harbourclean-up	+100
FRENCH	
#ramassagededechets	11497*
#ramassagededéchets	1658*
#ramassagededechet	1172*
#nettoyageplage	991
#ramassagedechets	978*
#collectededechets	929*
#collectededéchets	871*
#nettoyageplages	47
#collectededechet	30
#nettoyage_plage	7

**hashtags not directly related to beach clean-up activities.*

Hashtags counts on 5th November 2024. Note that in some cases the platform provided exact counts, and in other cases and approach. In Spain, Instagram media posts are published in Spanish, Galician and Catalonian languages; Euskera language is minoritarian in social platforms. English hashtags are widely present in any country. Therefore, many posts in Portugal, Spain or France may contain English hashtags.

During the filtering process, only events carried by volunteers were retained.

2.3. Other sources of beach clean-up initiatives

A multi-source search and data extraction strategy was implemented to compile a dataset addressing both the economic costs of beach cleaning and the scale of voluntary clean-up initiatives.

The search for cost data involved a systematic review of published literature, institutional studies and technical reports from international programmes, environmental agencies, and non-governmental organisations (NGOs). These sources provided quantitative estimates of beach cleaning costs across different countries, regions, and cleaning methods. In addition, project deliverables and public datasets from EU-funded initiatives and marine litter monitoring programmes were also reviewed. All figures were harmonised into a common format, standardising cost metrics to facilitate direct comparison across sources.

The search for voluntary clean-up initiatives drew on a combination of institutional platforms, campaign organisers, and NGO reporting. National databases such as MARNOBA provided detailed records of clean-up events in Spain, while international campaigns — including Ocean Conservancy's International Coastal Clean-up or Surfrider Foundation Europe's Ocean Initiatives — supplied further data.

Extracted information included the number of initiatives, participants, locations, and waste collected (by item counts or weight). When possible, averages such as participants per event were calculated. Regional distribution was recorded at both country and subnational level. The final dataset integrates all sources, enabling the production of summary statistics and allowing for the estimation of potential cost savings by applying harmonised cost metrics to the recorded voluntary activity.

3. RESULTS

3.1 Beach clean-up initiatives published in X

After developing the algorithm for capturing X data and adapting it to the countries of the Atlantic Area (Ireland, Spain, France, and Portugal), its effectiveness was evaluated. The basic X API access account only allows data from the previous week to be collected, so it is necessary to run the algorithm for a certain amount of time to capture a sufficient

amount of data. In this case, data was collected jointly for the four countries over a period of three months, from July 16th to October 15th, 2025.

Table 3 shows the number of posts retained by the algorithm, the number of initiatives that were ultimately valid and geolocated after human supervision, as well as other data of interest such as coastline length, number of beaches, and number of social media users.

TABLE 3: Number of tweets, coastal and population characteristics

Country	Number of posts captured by the algorithm	Number of beach clean-up initiatives after supervision	Coastline length (km)[(1)]	Number of beaches	Active users in X (% population)	Active users in Instagram (% population)
France	78	15	7330	2074 [(2)]	7.8M (11.4%)	29.1M (43.4%)
Ireland	1777	10	6437	~300 (153 recognized bathing waters) [(3)]	1.8M (33.4%)	2.64M (51.3%)
Portugal	51	13	2830	664 [(4)]	2.01M (18.8%)	5.3M (51.7%)
Spain	982	125	7268	3551 [(5)]	10.4M (21.3 %)	26.9M (55.8%)

[(1)] Data from the World Resources Institute, based on data calculated in 2000 from the World Vector Shoreline, United States Defense Mapping Agency, 1989. [(2)] Bath sites at sea. Data from eaufrance. [(3)] Data from the Environmental Protection Agency. [(4)] Including Azores and Madeira arquipelagos. Data from Instituto Nacional de Estatística. [(5)] Data from datos.gob.es. Monthly active users in social network for 2024 were obtained from Statista.

The number of tweets captured by the algorithm with the potential to mention a beach clean-up initiative in Ireland is the highest in comparison with the other countries of the Atlantic Area, with a total of 1,777 tweets. This is mainly because the tweets are in English and therefore have a wider global reach. It should be noted that the algorithm retains tweets that it has not been able to geolocate automatically for subsequent monitoring. This involves a time-consuming human post-processing. Despite the large number of potential tweets, only 10 tweets were finally associated with citizen initiatives in Ireland: one tweet was automatically geolocated in Lordship, County Louth, and the rest were obtained after monitoring the location declared by users in their

account profile. Another 14 tweets referred to a general call to action under the hashtag #BigBeachClean24, a clean-up campaign held between September 20th and 22nd.

The second largest set of tweets captured by the algorithm corresponds to those actions with the potential to take place in Spain. In this case, 125 initiatives were finally geolocated during this 3-months period, representing 12.8% of the initial set of potential tweets. France follows up, with 78 potential tweets that became 15 initiatives after supervision. In the case of Portugal, only 13 initiatives were collected from an initial set of 51 tweets, although the ratio is better than in the case of France. Many of these initiatives are accompanied by the hashtag #EUBeachCleanup 2024, a worldwide awareness raising campaign against marine litter, organized by the European Union in partnership with the United Nations.

Although the problem of capturing tweets from countries outside our target is greater when we use the English language (in the case of Ireland), France and Portugal also present a large proportion of the tweets from former colonies and overseas territories, which requires eliminating them in the supervision process.

Spain, by contrast, presents the stronger response to the algorithm, with the highest number of initiatives collected in these three months, as well as the best ratio in terms of both coastline length and number of beaches. This outcome may be due to a higher number of users on X in Spain than in other countries, with almost 10.4 million active users per month. However, it is important to mention that there may be a potential bias in the algorithm's design, as the authors are Spanish-speaking, allowing them to create better filters that exclude tweets that may be unrelated in the early stages.

The methodology tested during these three months shows very uneven performance between countries. Although the results obtained are optimal in Spain, they prevent further comparisons among countries.

3.2 Beach clean-up initiatives published in Instagram

Instagram has a higher degree of penetration in society than X, as can be seen in Table 3. The most important difference with respect to X in terms of the aim of this study is that searches can only be performed by hashtag, and not by free text as in X. This limits searches to a preselected list of hashtags and prevents straight comparison between these two social networks. Instagram has also a more limiting API's terms of use. However, it is possible to collect posts over a longer period, so posts from the whole 2024 were retrieved.

Table 4 shows the number of posts on Instagram published during 2024, sorted by the selected hashtag (see the methodology section). From 5,100 posts, 37 are repeated because they include several hashtags from the preselected list. It is noteworthy that the number of posts for some hashtags appears to be in the hundreds (e.g., #beachclean, #beachclean-up, #limpezadepraias), while others show the exact number (#beach_clean), which raises questions about how Instagram returns results. The same issue is reported in Table 3.

The hashtag #limpiezadeplayas is the most frequent (500 posts), followed by #limpiezadeplaya (490). In these posts, the geolocated initiatives take place mainly in Spain. The third most frequent hashtag is #limpezapraia (422), which involved 18 initiatives in Portugal and 7 in Spain. This hashtag is the same in Portuguese and Galician (a co-official language in the region of Galicia, NW Spain), which explains its appearance in posts published in both countries.

The hashtag #LimpezaSimultâneaDePraias (*simultaneous beach clean-up*) allows several locations to be recognized in a single post, making it highly effective for mapping citizen initiatives of this type in Spain.

In France, the hashtag #nettoyagedeplage is highly effective for locating volunteer clean-ups.

In Portugal, the hashtags #limpezapraia and #EUBeachCleanup are the ones most associated with posts with geolocated events.

If we subset our data to the 3-month period (16th July to 15th October 2025) to establish comparison with our testing of the X algorithm, we get 36 initiatives in Spain (125 in X), 14 in Portugal (13 in X), 9 in France (15 in X) and only 1 in Ireland (10 in X).

The results show very uneven performance between countries, with higher performance in Spain than in the other countries, both in X and Instagram. In Portugal, results are similar in both social networks. In contrast, the number of initiatives in Ireland and France are low, and the performance is even lower in Instagram than in X. However, this result clearly underestimates the actual situation, as occurs e.g. in the case of Ireland. There are in fact numerous initiatives in the country, such as An Taisce's Clean Coasts programme, a year-round initiative that engages more than 2,000 coastal communities across Ireland, primarily in beach clean-ups. This programme actively promotes its activities on social media, although many of the communities involved rely on Facebook as their main communication channel, reflecting the older demographic of the volunteers. On Instagram, relevant posts also exist, but they often use different hashtags from those preselected in our search strategy (for example, #CleanCoasts, #InternationalCoastalCleanup or #BigBeachClean). This divergence in hashtag usage

helps to explain why Irish initiatives were not fully captured by our dataset, despite their significant presence on the ground.

Spain and France both show a comparable amount of monthly active users in Instagram, with 26.9M and 29.1M people (see Table 3). However, the amount of people living in coastal municipalities highly differ between both countries. In France, 8M people live in coastal municipalities, ~12% of metropolitan population (Ministères Aménagement du Territoire Transition écologique, 2021). In contrast, in Spain, approximately 60% of the population resides in coastal regions, totalling around 27.7M people. This concentration is higher than the average for the European Union, where only 42% of the population lives in coastal areas, according to Eurostat www.coastal-management.eu. This could be an important factor affecting the number of posts published about marine and coastal topics and could explain the higher performance in Spain in both X and Instagram.

Another key factor is the way in which citizens interact with their natural environment, which is determined by the number of people living in coastal areas, the length of the coastline, the number of beaches, and also the leisure activities carried out in these spaces.

Spain has the highest number of beaches among the countries considered and also has a higher percentage of the population living in coastal areas. Additionally, favourable weather conditions encourage frequent use of beaches for leisure throughout the year. This could be a plausible explanation to the good performance observed.

In contrast, mainland Portugal has a lower number of beaches compared with Spain. Many of its beaches have extensions of up to several kilometres, and they are very exposed to sea conditions, making them harder to organize beach clean-up initiatives.

TABLE 4: Posts and beach clean-up initiatives in Instagram during 2024

Hashtags	Posts' counts	Geolocated and valid				
		All	ES	FR	IE	PT
beach_clean	156	0	0	0	0	0
beach_cleanup	155	0	0	0	0	0
beachclean	200	1	1	0	0	0
beachcleanup	200	4	1	3	0	0
beachcleanupchallenge	164	0	0	0	0	0
beachcleanupday	200	0	0	0	0	0

cleanthebeach	200	2	1	0	1	0
coastalcleanup	200	1	1	0	0	0
coastclean	187	0	0	0	0	0
EUBeachCleanup	200	18	4	0	1	13
harbourcleanup	141	0	0	0	0	0
limpezadepraia	192	0	0	0	0	0
limpezadepraias	200	11	9	0	0	2
limpezapraia	422	25	7	0	0	18
limpezapraias	128	3	3	0	0	0
LimpezaSimultáneadePraias*	12	34*	34*	0	0	0
Limpiezacostera	365	8	8	0	0	0
Limpiezadeplaya	490	29	29	0	0	0
Limpiezadeplayas	500	8	8	0	0	0
Limpiezalitoral	42	8	8	0	0	0
Limpiezaplayas	200	22	22	0	0	0
Netejadeplatges	146	9	9	0	0	0
Nettoyagedeplage	200	29	1	28	0	0
Playalimpia	200	5	5	0	0	0
TOTAL	**5100**	**217**	**151**	**31**	**2**	**33**

**Note that whereas only 12 posts were obtained with the hashtag #LimpezaSimultáneadePraias (simultaneous beach cleaning), these made possible to identify a total of 34 events on different beaches and/or dates.*

3.3 The particular case of Spain

The above results have demonstrated how social media can be used to gather a large number of beach clean-up initiatives in Spain. That is why in this section we focus on this country as a case study with the specific objectives of mapping citizen beach clean-up initiatives both spatially and temporally, gaining a better understanding of how these initiatives are publicized on social media, and comparing them with the number of initiatives compiled by environmental organizations, both governmental and non-governmental.

After applying the methodologies described above and extending the study to the entire year of 2024, a total of 487 volunteer clean-up initiatives were identified. This figure corresponds exclusively to those initiatives that could be geolocated. With this aim, initiatives advertised through multiple posts had to be counted appropriately. In other cases, a single post referred to multiple beach clean-ups. 68.9% of these initiatives were retrieved from social network X, with the remainder coming from Instagram (Figure 2).

Figure 2 shows the spatial distribution of the events, with the largest number of them in the region of Galicia. In this region, we also observe many initiatives come from Instagram, while X has a clear predominance in regions such as the Principality of Asturias, Cantabria, and the Valencian Community.

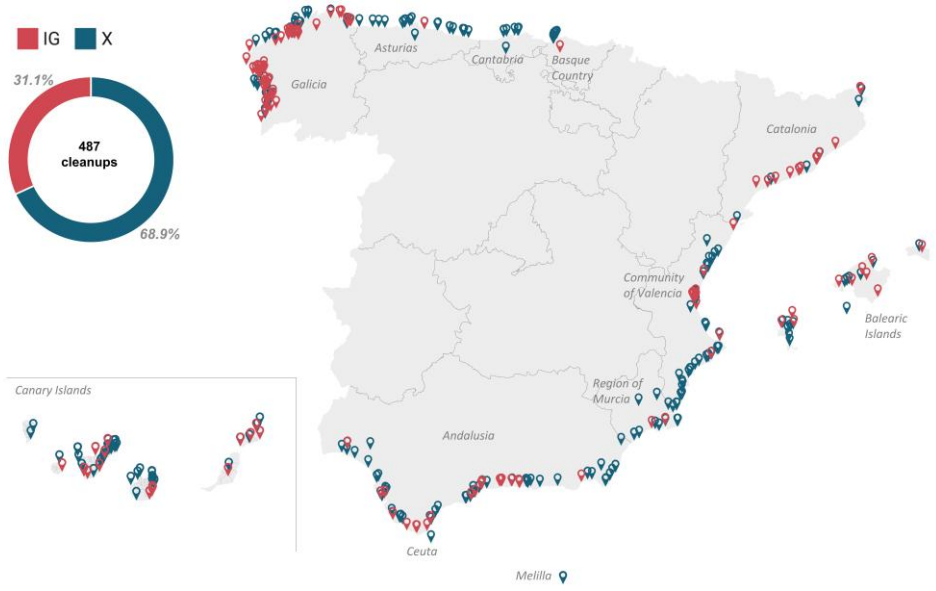

FIGURE 2: Spatial distribution of beach clean-up activities carried by volunteers in Spain during 2024, obtained from the social networks X and Instagram.

Table 5 shows the distribution by autonomous community. Although Galicia is the region with the highest count of beach clean-ups, as mentioned above, the Valencian Community is the region with the highest ratio of initiatives per length of coastline. However, if we consider the number of beaches, Andalusia is the region with the highest ratio.

In the autonomous cities of Ceuta and Melilla, only 2 and 4 beach clean-ups have been recorded, respectively. However, these regions have a very small number of beaches, only 16 and 8. Hence, Melilla, with only 8 km of coastline, has the highest ratio of initiatives per coastal length.

TABLE 5: Beach clean-ups by coastal regions in Spain during 2024

Region	Coastal length	Number of beaches	Inhabitants in coastal municipalities	Clean-ups	Ratio Clean-ups/ coast	Ratio Clean-ups/ beaches
Andalusia	945	402	3,310,122	90	0.0952	0.2239
Asturias	401	293	497,631	27	0.0673	0.0921
Balearic Islands	1428	340	1,231,768	23	0.0161	0.0676
Basque Country	246	67	967,681	9	0.0366	0.1343
Canary Islands	1583	589	2,238,754	84	0.0531	0.1426
Cantabria	284	93	448,921	8	0.0281	0.0860
Catalonia	699	477	4,589,663	26	0.0372	0.0545
Community of Valencia	518	333	3,542,636	66	0.1274	0.1982
Galicia	1498	861	1,601,411	129	0.0861	0.1498
Region of Murcia	274	192	494,182	19	0.0693	0.0989
Ceuta	20	16	83,179	2	0.1000	0.1250
Melilla	8	8	85,985	4	0.5000	0.500

**Inhabitants in coastal municipalities have been calculated taking into account the annual population census by region as of 1st January 2024 (Insituto Nacional de Estadística, 2024), and the ratio of coastal population in each region in 2020 (MAPA, 2022). All inhabitants of Ceuta, Melilla, Balearic Islands and Canary Islands are considered coastal population.*

Figure 3 shows the temporal evolution of beach clean-up events. As can be observed, there is a peak in January corresponding to beach clean-ups organized mainly in response to the spill of pellets from the Toconao ship. This spill resulted from the loss of six containers from this cargo off the coast of Portugal, spilling 26.3 tons of plastic pellets into the sea, in addition to another five containers with tires and plastic film. Galician beaches became the epicentre of this disaster and are therefore where the greatest

number of beach clean-up events took place under numerous initiatives promoted through social media.

During February and March, there were few events, but they gradually increased again from April onwards, reaching another peak in June. These events took place during the spring, with the arrival of favourable weather conditions, which encourages such initiatives and allows the beaches to be cleaned of the debris and floating objects deposited by winter storms. During the summer, the number of initiatives decline again, as these are months when beaches are heavily used for tourism and leisure, and it is mainly coastal municipalities that take care of beach cleaning, often using mechanical methods. After the summer season, beach clean-up initiatives regain popularity in September and October, coinciding in large part with the celebration of the #EUBeachCleanup — a worldwide awareness-raising campaign against marine litter, organized by the European Union in partnership with the United Nations and the Smurfs. World Clean-up Day 2024 was on Friday, September 20th, but any event organized between June 10th and October 31st counts toward the campaign. It is also during October that the highest number of Instagram events has been recorded compared to X, something that did not happen during the rest of the year, indicating the importance of this social network for publicizing this event worldwide. In fact, the hashtag #EUBeachCleanup is one of the most frequently used to announce these initiatives on Instagram (Table 4).

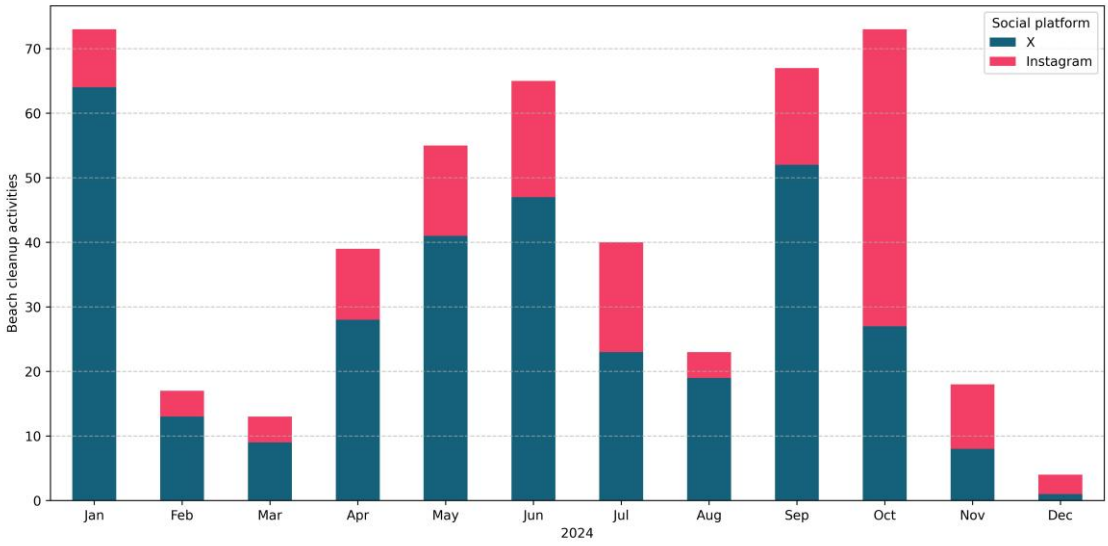

FIGURE 3: Temporal (monthly) distribution of beach clean-up activities carried by volunteers in Spain during 2024, obtained from the social networks X and Instagram.

According to published data, during 2024, 367 events were organised in 53 countries under the #EUBeachCleanup initiative, involving 12,270 volunteers. In Spain, this

initiative comprised 25 registered activities, as personally communicated the #EUBeachCleanup secretariat to the authors (mail on 5 Sep 2025).

Other international campaigns such as the International Coastal Clean-Up coordinated by Ocean Conservancy have also played a significant role, with Spanish participation involving several thousand volunteers across more than 100 sites during 2024 (see Table 6).

In Spain, large-scale initiatives such as Proyecto Mares Circulares led by the Coca-Cola® Foundation and partners, and the Proyecto *LIBERA – 1 m² por las playas y los mares* mobilised thousands of volunteers across multiple regions. The first one engaged over 9,700 volunteers across 169 beaches, 22 marine reserves, and eight underwater areas, removing more than 376 tonnes of litter. Proyecto LIBERA gathered more than 5,700 participants at 360 coastal sites, recording over 108,000 litter items.

Other two prominent citizen-led initiatives carried out in Spain shows distinct patterns of engagement. Innoceana events had an average of 8.9 volunteers per clean-up, typically representing smaller, more localised actions. In contrast, Good Karma clean-ups averaged 35.7 volunteers per event, reflecting larger-scale operations with a broader reach. Both models contribute significantly to litter removal and environmental awareness, highlighting the diversity of formats through which voluntary action can occur. Despite this wide range, and based on data from Table 6, we can assume an average of 24 volunteers per activity.

TABLE 6: Beach clean-up activities in Spain during 2024 reported by several organizers

Initiative / Project	Geogr. scope	Organiser	No. of events	Participants		Notes
				Total	Average per event	
Ambiente Europeo & International Coastal Clean-up España	Spain	Ambiente Europeo	105	2,589	24.66	111,640 items, 3,085 kg

Ocean Conservancy & International Coastal Clean-up	Spain	Ocean Conservancy		4160		287,248 items, 7,518.42 kg
Proyecto Mares Circulares	Spain + Portugal	Coca-Cola Foundation, Chelonia, Ecomar, Vertidos Cero, Liga para a Protecção da Natureza	169 beaches + 22 reserves + 8 underwater areas	9,731	n/a	376 t removed
MARNOBA Vertidos Cero (viewer)	Spain		695			662250 items
Proyecto LIBERA – 1 m² por las playas y los mares	Spain	SEO/BirdLife & Ecoembes	360	5,742	15.95	108,887 items
Innoceana	Spain (Canary Islands)	Innoceana	18	160	8.3	
Cleanup with Good Karma	Spain	Good Karma	32	1142	35.7	
Plan de Conservación Territorial– ON (PLANCTON)	Spain (Galicia)	Obra social Abanca Afundación	n/a	1,000 + 200 boats	n/a	43 t
Campaña “Limpiemos el Mar”	Spain (Mediterranean & Atlantic)	Ecologistas en Acción	5	150	30	>0.5 t
Iniciative 'La Costa y Tú'	Spain (Bizkaia y Gipuzkoa)	Surfrider Foundation		900		>7000 butts

Notes:

- "n/a" indicates data not provided in the original records.
- Where possible, averages are calculated directly from the number of participants and events.
- The table merges duplicate references to the same initiative while retaining the most complete dataset for each.

After inspection of the previous 487 activities obtained from social network data, we have identified 22 events related to the campaigns mentioned in Table 6. Therefore, the methodology presented in this study has allowed us to identify 465 events not recorded in any of the databases and reports considered.

3.4 Beach cleaning costs and potential savings from voluntary action

At the EU level, the average estimated costs per km of beach cleaning is €40,000/year (Commission et al., 2018). However, it is difficult to obtain a precise and up-to-date estimate, because data about beach clean-ups are scarce and heterogenous among countries, and sometimes with studies that are more than 15 years old.

For example, in the Netherlands and Belgium, cost estimates range from low-cost interventions, such as those carried out by volunteers (≈ €0.10/m of shoreline) (Doomen et al., 2009; Reinhard et al., 2012), to mechanised operations (ranging from 36 to 45 €/ha) (Doomen et al., 2009), and exceeding €97,000/km/year (Mouat et al., 2010; Newman et al., 2015). Municipalities spend an average of € 25/h on manual cleaning, or about € 3.600 ha/year, while for mechanical cleaning, the cost is approximately € 5.775,76 ha/year (Doomen et al., 2009).

In United Kingdom, the estimated costs present wide range from, €171 to €82,000/km/year and average of €7,000–7,300/km/year. In this country, it has been estimated that each volunteer contributes the equivalent of €16.23/year of their time each year on average to remove litter removal, with a total value exceeding €131,000 for two of the largest clean up national schemes (MCS Beachwatch and KSB National Spring Clean) (Mouat et al., 2010). In Ireland and Portugal, the estimated costs are €11,244–12,850/km/year and €8,278–31,768/km/year; and Sweden, ≈ €214–4,580/km/year (Mouat et al., 2010; Newman et al., 2015). Higher costs correlated with more intense cleaning operations on small areas of coastline, particularly in tourist areas

(Newman et al., 2015). Thus, in Germany, SPA resorts paid every year up to 38 €/m/year for beach cleaning and biomass disposal (Mossbauer et al., 2012).

In Spain, reported values range between €38,189.75 and €87,500/km/year depending on the degree of pollution and maintenance frequency (Mouat et al., 2010). In a case study in the Gulf of Cádiz (South of Spain), the computed mean cost was €50,376/km/year, with a cost per weight collected varying between €0.98 and €2.61/kg (mean €1.34/kg) and an estimated productivity of 27.6 kg/person/hour (Cruz et al., 2020).

Taking into account the most updated costs of Cruz et al. (2020), based on data collection from 2010 to 2012, and assuming an average of 24 volunteers per event when no data are available, we get 41,112 volunteers in Spain involved in the campaigns of Table 6. Assuming also a duration of 2h of litter collection per activity, and using the already mentioned productivity of 27.6 kg/person/hour, the total costs are roughly 3M€. If we add the initiatives gathered in X and Instagram, the costs ascend to 3,866,603€. If we assume an interannual increase of labour costs in Spain of 3.8% (data obtained from Instituto Nacional de Estadística, www.ine.es), in 2025 we estimate a total cost of 4.57M€.

4. CONCLUSIONS

This study presents a novel methodology to retrieve citizen beach clean-up events trough social network data mining. The methodology tested showed very uneven performance among the countries of the Atlantic Area, which prevents further comparison. However, the method has proven to be highly effective in Spain. This may be due to a higher number of users on X than in other countries, a larger population living in coastal areas and the predominant weather conditions, which lead to Spanish beaches becoming centres of recreation and leisure.

The test of this methodology in a case study in Spain during 2024 allowed the retrieval of 487 volunteer beach clean-up events. Two thirds of these events were retrieved through the X social network, with a peak during January related to the Toconao pellets' spill. After comparison with other national and international campaigns, 465 unique events were identified as not reported by other organizations in Spain during this period. Previous economic studies reveal a high variability in beach cleaning costs, with values depending on the region, the intensity of interventions and the cleaning method. Furthermore, we must bear in mind that the estimates are produced in very different years, so the evolution of the consumer price index in each country would have to be taken into account. However, taking into account the most recent studies in 2020, we

can estimate a total costs of beach clean-up initiatives by volunteers in Spain of 4.57M€, with 21% of these costs corresponding to initiatives mined from social networks and not reported in other databases.

Our data demonstrate that voluntary action is a widespread and coordinated phenomenon across the Atlantic Area region. This multi-country mobilisation not only enhances the scale of litter removal but also represents a considerable social capital investment, with citizens dedicating time and resources to coastal protection. These data enable an assessment of both the geographical distribution and intensity of voluntary engagement, providing a basis for estimating the economic value of voluntary clean-ups when compared with the costs of equivalent operations carried out by contracted services.

Importantly, the case of Ireland highlights the need to refine social media monitoring approaches by considering both country-specific platform preferences and the socio-demographic characteristics of the communities involved. The reliance on certain platforms, combined with the use of localised hashtags or terminology, can significantly affect detection and bias cross-country comparisons. To achieve a more accurate representation of volunteer initiatives, future research should therefore tailor search strategies to the digital communication practices and the degree of adoption of each social network in different regional contexts.

5. REFERENCES

Arts, K., van der Wal, R., Adams, W.M., 2015. Digital technology and the conservation of nature. Ambio 44, 661–673. https://doi.org/10.1007/s13280-015-0705-1

Barker, J.L.P., Macleod, C.J.A., 2019. Development of a national-scale real-time Twitter data mining pipeline for social geodata on the potential impacts of flooding on communities. Environmental Modelling & Software 115, 213–227. https://doi.org/10.1016/j.envsoft.2018.11.013

Budenz, A., Klassen, A., Purtle, J., Yom Tov, E., Yudell, M., Massey, P., 2020. Mental illness and bipolar disorder on Twitter: implications for stigma and social support. J Ment Health 29, 191–199. https://doi.org/10.1080/09638237.2019.1677878

Commission, E., Environment, D.-G. for, Eunomia, ICF, 2018. Assessment of measures to reduce marine litter from single use plastics – Final report and annex. Publications Office. https://doi.org/doi/10.2779/500175

Cooper, M.W., Di Minin, E., Hausmann, A., Qin, S., Schwartz, A.J., Correia, R.A., 2019. Developing a global indicator for Aichi Target 1 by merging online data sources to measure biodiversity awareness and engagement. Biological Conservation 230, 29–36. https://doi.org/10.1016/j.biocon.2018.12.004

Correia, R.A., Jepson, P., Malhado, A.C.M., Ladle, R.J., 2017. Internet scientific name frequency as an indicator of cultural salience of biodiversity. Ecological Indicators 78, 549–555. https://doi.org/10.1016/j.ecolind.2017.03.052

Cruz, C.J., Muñoz-Perez, J.J., Carrasco-Braganza, M.I., Poullet, P., Lopez-Garcia, P., Contreras, A., Silva, R., 2020. Beach cleaning costs. Ocean & Coastal Management 188, 105118. https://doi.org/10.1016/j.ocecoaman.2020.105118

Dang, T., Nguyen, N.V.T., Pham, V., 2018. HealthTvizer: Exploring Health Awareness in Twitter Data through Coordinated Multiple Views, in: 2018 IEEE International Conference on Big Data (Big Data). Presented at the 2018 IEEE International Conference on Big Data (Big Data), pp. 3647–3655. https://doi.org/10.1109/BigData.2018.8622445

Di Minin, E., Tenkanen, H., Toivonen, T., 2015. Prospects and challenges for social media data in conservation science. Front. Environ. Sci. 3. https://doi.org/10.3389/fenvs.2015.00063

Doomen, A., van Vliet, J., Musters, I., Drost, J., van Bezooijen, G., 2009. Duurzaam strandbeheer. Voor een schoon en natuurlijk strand (No. 275016.rsd.R003). Middelburg.

Gliozzo, G., Pettorelli, N., Haklay, M. (Muki), 2016. Using crowdsourced imagery to detect cultural ecosystem services: a case study in South Wales, UK. Ecology and Society 21.

Hausmann, A., Toivonen, T.K., Slotow, R., Tenkanen, H.T.O., Moilanen, A.J., Heikinheimo, V.V., Di Minin, E., 2018. Social Media Data Can Be Used to Understand Tourists' Preferences for Nature-Based Experiences in Protected Areas. Conservation Letters 11. https://doi.org/10.1111/conl.12343

Heikinheimo, V., Tenkanen, H., Hiippala, T., Toivonen, T., 2018. Digital imaginations of National Parks in different Social Media: a data exploration., in: Westerholt, R., Mocnik, F.-B., Zipf, A. (Eds.), Proceedings of the 1st Workshop on Platial Analysis (PLATIAL'18). Heidelberg university, Heidelberg, pp. 45–52. https://doi.org/10.5281/zenodo.1472745

Instituto Nacional de Estadística, 2024. Censo Anual de Población. Variables de residencia, educativas y laborales de 2024 y 2023. Data published on 19 December 2024.

Ladle, R.J., Correia, R.A., Do, Y., Joo, G.-J., Malhado, A.C., Proulx, R., Roberge, J.-M., Jepson, P., 2016. Conservation culturomics. Frontiers in Ecology and the Environment 14, 269–275. https://doi.org/10.1002/fee.1260

Makita, M., Mas-Bleda, A., Morris, S., Thelwall, M., 2021. Mental Health Discourses on Twitter during Mental Health Awareness Week. Issues in Mental Health Nursing 42, 437–450. https://doi.org/10.1080/01612840.2020.1814914

MAPA, 2022. Demografía de la población costera en España, 2020 (No. 5), Análisis y Prospectiva – serie Pesca. S.G. Análisis, coordinación y estadística. Ministerio de agricultura, pesca y alimentación.

Martínez-Rojas, M., Pardo-Ferreira, M. del C., Rubio-Romero, J.C., 2018. Twitter as a tool for the management and analysis of emergency situations: A systematic literature review. International Journal of Information Management 43, 196–208. https://doi.org/10.1016/j.ijinfomgt.2018.07.008

Milani, E., Weitkamp, E., Webb, P., 2020. The Visual Vaccine Debate on Twitter: A Social Network Analysis. Media and Communication 8, 364–375. https://doi.org/10.17645/mac.v8i2.2847

Ministères Aménagement du Territoire Transition écologique, 2021. Coastline and Marine Environment in Metropolitan France. Extract from France's 2021 Environmental Performance Review. France.

Mossbauer, M., Haller, I., Dahlke, S., Schernewski, G., 2012. Management of stranded eelgrass and macroalgae along the German Baltic coastline. Ocean & Coastal Management 57, 1–9. https://doi.org/10.1016/j.ocecoaman.2011.10.012

Mouat, J., Lozano, R.L., Bateson, H., 2010. Economic Impacts of marine litter. KIMO International.

Newman, S., Watkins, E., Farmer, A., Brink, P. ten, Schweitzer, J.-P., 2015. The Economics of Marine Litter, in: Bergmann, M., Gutow, L., Klages, M. (Eds.), Marine Anthropogenic Litter. Springer International Publishing, Cham, pp. 367–394. https://doi.org/10.1007/978-3-319-16510-3_14

Otero, P., Gago, J., Quintas, P., Incera, M., Pérez, P., Cabrero, Á., González-Nuevo, G., 2021. Tweeting about marine plastic pollution: a one-year exploratory analysis. Presented at the "One Integrated Marine Debris Observing System", an official satellite activity of the UN Ocean Decade 3rd laboratory, On-line. https://doi.org/10.13140/RG.2.2.16958.20807

Otero, P., Menéndez-Blázquez, J., March, D., 2025. Challenges of passive citizen science in ecology within a shifting social media landscape. Ecological Informatics 90, 103278. https://doi.org/10.1016/j.ecoinf.2025.103278

Otero, P., Velasco, E., Valeiras, J., 2024. Surveillance of coastal biodiversity through social network monitoring. Ecological Informatics 80, 102515. https://doi.org/10.1016/j.ecoinf.2024.102515

Plackett, R., Kaushal, A., Kassianos, A.P., Cross, A., Lewins, D., Sheringham, J., Waller, J., von Wagner, C., 2020. Use of Social Media to Promote Cancer Screening and Early Diagnosis: Scoping Review. J Med Internet Res 22, e21582. https://doi.org/10.2196/21582

Poorthuis, A., Zook, M., 2017. Making Big Data Small: Strategies to Expand Urban and Geographical Research Using Social Media. Journal of Urban Technology 24, 115–135. https://doi.org/10.1080/10630732.2017.1335153

Reinhard, S., de Blaeij, A., Bogaardt, M.-J., Gaaf, A., Leopold, M., Scholl, M., Slijkerman, D., an Strietman, W., van der Wielen, P., 2012. Cost-effectiveness and cost-benefitanalysis for the MSFD. Framework for the Netherlands (No. LEI report 2011-036). LEI Wageningen UR, LEI, part of Wageningen UR, The Hague.

Richards, D.R., Friess, D.A., 2015. A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. Ecological Indicators 53, 187–195. https://doi.org/10.1016/j.ecolind.2015.01.034

Richards, D.R., Tunçer, B., 2018. Using image recognition to automate assessment of cultural ecosystem services from social media photographs. Ecosystem Services, Assessment and Valuation of Recreational Ecosystem Services 31, 318–325. https://doi.org/10.1016/j.ecoser.2017.09.004

Roll, U., Mittermeier, J.C., Diaz, G.I., Novosolov, M., Feldman, A., Itescu, Y., Meiri, S., Grenyer, R., 2016. Using Wikipedia page views to explore the cultural importance of global reptiles. Biological Conservation, Advancing reptile conservation: Addressing knowledge gaps and mitigating key drivers of extinction risk 204, 42–50. https://doi.org/10.1016/j.biocon.2016.03.037

Ruiz-Frau, A., Ospina-Alvarez, A., Villasante, S., Pita, P., Maya-Jariego, I., de Juan, S., 2020. Using graph theory and social media data to assess cultural ecosystem services in coastal areas: Method development and application. Ecosystem Services 45, 101176. https://doi.org/10.1016/j.ecoser.2020.101176

Shimkhada, R., Attai, D., Scheitler, A.J., Babey, S., Glenn, B., Ponce, N., 2021. Using a Twitter Chat to Rapidly Identify Barriers and Policy Solutions for Metastatic Breast Cancer Care: Qualitative Study. JMIR Public Health and Surveillance 7, e23178. https://doi.org/10.2196/23178

Teoh, D., Shaikh, R., Vogel, R.I., Zoellner, T., Carson, L., Kulasingam, S., Lou, E., 2018. A Cross-Sectional Review of Cervical Cancer Messages on Twitter during Cervical Cancer Awareness Month. J Low Genit Tract Dis 22, 8–12. https://doi.org/10.1097/LGT.0000000000000363

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., Di Minin, E., 2019. Social media data for conservation science: A methodological overview. Biological Conservation 233, 298–315. https://doi.org/10.1016/j.biocon.2019.01.023

Vraga, E.K., Stefanidis, A., Lamprianidis, G., Croitoru, A., Crooks, A.T., Delamater, P.L., PFOSER, D., Radzikowski, J.R., Jacobsen, K.H., 2018. Cancer and Social Media: A Comparison of Traffic about Breast Cancer, Prostate Cancer, and Other Reproductive Cancers on Twitter and Instagram. Journal of Health Communication 23, 181–189. https://doi.org/10.1080/10810730.2017.1421730

Yoosefi Nejad, M., Delghandi, M.S., Bali, A.O., Hosseinzadeh, M., 2019. Using Twitter to raise the profile of childhood cancer awareness month. Netw Model Anal Health Inform Bioinforma 9, 3. https://doi.org/10.1007/s13721-019-0206-4

Zhou, S., Kan, P., Huang, Q., Silbernagel, J., 2021. A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura. Journal of Information Science 01655515211007724. https://doi.org/10.1177/01655515211007724

6. SUPPLEMENTARY MATERIAL

TABLE A. Examples of posts in X social network captured by the algorithm and the result after human supervision. Posts have been deliberately anonymized.				
DATE	**TEXT**	**COMMENT**	**LANGUAGE**	**VALID**
10/08/2024	Plage de #Bargny #Ndiolmane #Nettoyage #Sensibilisation et #distribution de #poubelles Notre plage est un trésor que nous partageons tous, et il est de notre responsabilité commune de la garder propre @crac02 en partenariat avec SOCOCIM vous invite ce dimanche 11 Août https://t.co/e4RjKooDnF	Not located by the algorithm. Post from Senegal.	French	NO
19/07/2024	Le Conseil Municipal des Jeunes vous invite à participer au nettoyage de la plage le 23 juillet et le 20 août. L'événement se déroulera de 9h30 à 11h30 sur la plage de Trouville-sur-Mer. ♻🌍 ☎ Pour plus d'informations, vous pouvez contacter le 06 76 36 43 36. https://t.co/raqa0RBxNx	Located by the algorithm	French	YES
18/09/2024	Dans le cadre de la Journée mondiale du World Clean-up day (Journée mondiale de nettoyage de la planète), le Seaquarium ouvrait le bal ce mercredi 18 septembre avec un nettoyage public sur la plage située devant l'institut marin. https://t.co/uySzkRVHOo https://t.co/CvQQ9bsQl1	Not located by the algorithm. Location know with additional info in the text and the user location	French	YES
20/09/2024	#WorldClean-upDay2024 92 points de ramassage à @Toulouse pour ce jour de nettoyage. Citoyens, assos, entreprises, agents, élus, tous unis pour une planète + propre. Merci aux centaines de Toulousains mobilisés à nos côtés avec le sourire. 🚮🌎 @gnadango https://t.co/xW8Sqh9VNe	Not located by the algorithm. Clean-ups in rivers and inland locations.	French	NO
20/09/2024	Almanarre	Located by the algorithm	French	YES
26/07/2024	Vessels in the water and crews patrolling beaches are performing clean-up as necessary on Nantucket,	Not located by the algorithm.	English	NO

	while overflight monitoring continues as we assist in the response to the damaged GE Vernova blade. Collections include incident debris and other refuse in the monitored areas. 1/2 https://t.co/u3dGSQYxVG	No citizen initiative.		
16/07/2024	Singapore oil spill beach clean-up progressing 'faster than expected' https://t.co/11pmI0A4sG https://t.co/jVxLk1fNAl	Not located by the algorithm. It is not a citizen initiative and it's in Singapore.	English	NO
18/07/2024	We 💙 our #Florida beaches! Morning #clean-up starts at 🕗 w/ #BocaRaton Save our Beaches #BocaSaveourBeaches https://t.co/yVL8G1WJci	Not located by the algorithm. It is a citizen initiative in USA	English	NO
10/09/2024	Beach clean-up live Ireland AM with all the TY students from St.Pauls & Clean Coasts crew 😊 📺 IE 🙏 @CleanCoasts	Not located by the algorithm. Located after supervision.	English	YES
13/09/2024	Join us for a walk and talk at Rossnowlagh Beach on the 14th September at 1pm, meeting at the Franciscan Friary! Register your attendance here: https://t.co/JPJ2d5kULt @NPWSIreland @HeritageHubIRE #Donegal #CleanCoasts #Clean-up https://t.co/y3K5BFmV6q	Located and with a hashtag about coastal clean-up. However, it is not a beach clean-up activity.	English	NO
02/08/2024	Only 3 days left until the #BigBeachClean24 kicks off, with over 600 clean-ups taking place be sure to follow steps to stay safe during the clean-up. For extra tips & guidance: https://t.co/3AjxWtUsae 🚴 🌊 #CleanCoasts #CitizenScience #Ireland @kia_ireland @OurOcean https://t.co/TTRjumABXe	Not located by the algorithm. No specific location. Call to action!	English	YES
19/07/2024	An organised beach clean-up is taking place in the dunes of Brittas Bay Beach on Monday 22nd July from 17.30-19.30. If you wish to help out on the evening, please email eao@wicklowcoco.ie to register. Let's help keep Brittas Bay beautiful!! #Wicklow #YourCouncil https://t.co/eBYxJjduK3	Not located by the algorithm. Located after supervision.	English	YES

05/09/2024	📣 Notice to Mariners No. 13 of 2024 🚸 In the interest of environmental protection and safety to persons, a clean-up of The League, Reen Pier, Castletownshend Harbour will be carried out by Cork County Council on the 24th of September 2024. ➡️ Read more: https://t.co/fVs0ENqxVD https://t.co/AWKC3TMEdJ	Not located by the algorithm. Council initiative.	English	NO
08/08/2024	Mais fotos da Nina Dobrev na 4ª edição anual da limpeza de praia do Shiseido Blue Project em Huntington Beach, Califórnia. [07/08] https://t.co/qSEPmOBQ6d	Not in Portugal	Portuguese	NO
22/07/2024	A manhã de sexta-feira foi bem passada, numa atividade das 👧👦🌿 crianças do CCCRAM (Moinho Sábio) e do CASM. Os mais jovens promoveram a Terceira Caminhada Ecológica, contribuindo para a limpeza da praia na Ponta dos Corvos. https://t.co/9t38l12Fhv	Not located by the algorithm. Located after supervision.	Portuguese	YES
09/10/2024	🌊 Junta-te a nós na #EUBeachCleanup 2024! 🌍✨ O #LAQV/REQUIMTE está a organizar uma ação de limpeza costeira na Praia de São Pedro de Maceda (Ovar) 🏖️🗓️ 19 de out. 10h-12H 📝 Inscreve-te até dia 17 de out através do QR code do flyer! Há a opção de partilha de 🚙! 🌍💙 https://t.co/Ls7D3wAVun	Located. Call to Action!	Portuguese	YES
20/07/2024	🏖️ La feina a l'estiu a les platges de #Dénia... 🧹 Neteja i desbrossament manuals de la zona del mirador de cetacis de les Rotes 👊🤌 --------- 🏖️ El trabajo en verano en las playas de #Dénia... 🧹 Limpieza y desbroce manuales de la zona del mirador de cetáceos de Les Rotes 👊🤌 https://t.co/2SjR0rlAIi	Not located by the algorithm. No citizen initiative.	Catalonian	NO
24/07/2024	San Fernando implica a la UCA en su recurso contra las restricciones en la limpieza de la playa de Camposoto https://t.co/UlnHXtJaLb https://t.co/PItjNgE4bV	Located by the algorithm. Not beach clean-up initiative.	Spanish	NO
11/10/2024	Muy orgullosos de haber participado en el #ProyectoLIBERA '1m2 por las	Located by the algorithm	Spanish	YES

	playas y los mares'. Los alumnos de 2º ESO realizaron una limpieza en la playa de Los Quebrantos. La actividad se enmarcó en el componente de #ServicioComoAcción del #ProgramaDeLosAñosIntermedios. #LIBERA1m2 https://t.co/KkR10NaiBX			