[Protocol] PubliCo. A new risk and crisis communication platform to bridge the gap between policy makers and the public in the context of the COVID-19 crisis

Giovanni Spitale

Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

Sonja Merten

Swiss Tropical and Public Health Institute, Basel, Switzerland

University of Basel, Basel, Switzerland

Kristen Jafflin

Swiss Tropical and Public Health Institute, Basel, Switzerland

University of Basel, Basel, Switzerland

Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

Bettina Schwind

Swiss Tropical and Public Health Institute, Basel, Switzerland

University of Basel, Basel, Switzerland

Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

Andrea Kaiser-Grolimund

Swiss Tropical and Public Health Institute, Basel, Switzerland

University of Basel, Basel, Switzerland

Nikola Biller-Andorno (corresponding Author)

Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

Abstract

Introduction

Since the end of 2019, COVID-19 has unfolded significant impact on citizens around the globe. While more severe measures are being reinstalled in order to curb flareups, it is likely that public adherence will gradually decrease, and that discontent will start to mount. Providing high-quality information and countering fake news is not enough. What is needed in addition is creating feedback loops: In order to continuously refine and adjust preventive measures and communication strategies, policy-makers and authorities need

information – preferably based on real-time data – on how the public receive the messages delivered, not only at a cognitive but also at an emotional and behavioural level.

Methods and Analysis

Against this backdrop, our project aims to develop a tool that helps tackle the "infodemic" manifested in the COVID-19 context, with a focus on a nuanced and in-depth understanding of public perception. We intend to foster effective and tailored risk and crisis communication as well as an assessment of the risks and benefits of prevention, containment and control measures, as their effectiveness will crucially depend on public trust and cooperation. The project adopts a trans-disciplinary multi-stakeholder approach including a participatory citizen science component. Methodologically, we propose a combination of literature and media review and analysis, and empirical studies using mixed methods. Building on real-time data and continuous data collection, our research results will be highly adaptable to the evolution of the current emergency (e.g., with respect to vaccine development).

Ethics and dissemination

PubliCo does not fall under the scope of the Swiss Human Rsearch Act. However, being conscious of the risks that such a tool could create in a non-democratic context, we meticulously self-assessed potential risks and took care to ensure data protection in the coding and structure of PubliCo.

Project results, interpretative briefs and intermediate datasets will be publicly available via the project website (www.publico.community) and via open science repositories (Zenodo).

Strengths and limitations of this study

- PubliCo is a new modular and flexible tool to provide bi-directional interaction between citizens and policy-makers for risk and crisis communication
- PubliCo relies on quantitative and qualitative data to provide a precise, timely and rich analysis of complex phenomena
- PubliCo is open and transparent by design
- Although important safeguards are put in place at a code level, in a non-democratic context it could be used to spy on people
- Communicating complex notions with moral implications (e.g. health risks) is a challenge.

Introduction

Background

Since the end of 2019, COVID-19 has unfolded significant impact on citizens around the globe. Beyond infections, disease and death, the global public has been exposed to policy measures that quickly escalated in scale and severity. Within weeks or even days, measures evolved from recommendations, such as frequent handwashing, to more disruptive interventions, including cancellations of social events, closure of schools, locked borders, a ban of gatherings of more than five people in public space, and a physical distance requirement of two meters. Public life and ways of socializing that were taken for granted have come to an abrupt halt.

In Switzerland measures have been moderate, in comparison to other countries, yet more drastic dispositions are conceivable and legally covered by the Swiss Epidemics Law, should the situation require them, such as a general curfew, mandatory testing or the use of mobile phone data for surveillance purposes. During the first wave (March to June 2020), the Swiss population has taken these measures fairly well, and encouraging signs of responsiveness and solidarity have become apparent. As the second wave unfolds, however, there is an increasingly controversial debate on issues such as contact tracing, visits in nursing homes, home office, and economic implications.

In general, it is to be expected that the exceptional circumstances will bring significant short-, mid- and long-term consequences, in social, economic and maybe cultural and political terms. Some issues have already sprung up, including social isolation of vulnerable groups, such as patients in hospitals and the elderly in nursing homes due to visiting bans, panic buying and stolen supplies, or instances of reprimanding others for their "irresponsible" behaviour. The gradual easing of containment measures has certainly relaxed growing frustration in parts of the population.

Nevertheless, while more severe measures are being reinstalled in order to curb flareups, it is likely that public adherence will gradually decrease, and that discontent will start to mount. First signs of resistance have already become apparent in the form of "anti-corona" demonstrations in several cities, gatherings of hundreds of persons celebrating the end of the lockdown or organized 'illegal' soccer games [1].

Information gaps

Amidst this complex situation, while the government is trying to steer through this crisis as cautiously as possible, the public is grappling with how to interpret what is happening. Existing literature suggests that effective health communication can help enhance positive outcomes of public policy [2,3]. Importantly, the exposure to focused health campaigns in the context of epidemics has proven as an efficient tool not only to increase epidemic-related knowledge, but also to foster the adoption of recommended health behaviours [4]. Communication is therefore key.

Despite the fact that international organizations, national governments, health authorities, scientific institutions and high-quality media are trying to tackle this task as responsibly and professionally as possible, all over Europe a flurry of newsfeeds, reports, tickers, trackers and live maps of questionable credibility exist across media platforms. These contents might be selectively echoed and reinforced by formal and informal opinion groups and influence public opinions in problematic ways, e.g., blaming of specific social and ethnic groups or decreasing willingness to adhere to rules set by authorities. Some media draw on dystopic pictures and morally loaded language, using war metaphors and reproaching those who voice doubts and criticism, which leads to polarization and an affectively charged debate producing strong counterreactions rather than factual and nuanced public deliberation [5]. The WHO has appropriately warned of an "infodemic" flooding the public with a mixture of reliable and false information [6]. WHO regional offices alongside other public health agencies work intensely on refuting myths regarding, e.g., false preventive measures and false cures, through fact checks of social media and writing responses[7].

Providing high-quality information and countering fake news, however, is not enough. What is needed in addition is creating feedback loops: In order to continuously refine and adjust preventive, control and containment measures and communication strategies, policy-makers and authorities need information – preferably based on real-time data – on how the public receive the messages delivered, not only at a cognitive but also at an emotional and behavioural level.

Understanding how mitigation measures are received by the population allows for an estimation of how effective they are likely going to be and may thus influence not only communication strategies but also policy choices [5,8]. It will also help us understand to what extent policy decisions and other options match with citizens' moral values and preferences regarding, e.g., the allocation of scarce medical resources (such as, for instance, a new vaccine), personal or digital contact tracing, or obligatory mask wearing [9]. Finally, policy-making will profit from understanding how the current situation is perceived by different parts of the population, in particularly by those who are vulnerable, such as frontline healthcare workers, the elderly, the chronically ill, or those in difficult economic circumstances. Since using a "one size fits all" approach as

mitigation measures in the context of epidemics has noticeable limitations, local and subgroup data are critically needed to deploy more efficient strategies [10].

So far, there has been mainly "one-way communication". This means, for instance, we do not know much about how different subgroups may differ in their understanding of the situation and their readiness to comply with policies, and how this is affected by their preferred information sources. Selective opinion polls or individual opinions cannot fill this gap. Also one-time online or phone surveys [11–13] encounter important limits in view of the rapidly evolving situation – they are resource-intense and limited in scope, their items are typically designed in a top-down way, they struggle with high non-response rates and provide snapshots rather than continuous monitoring [9,14]. Consequently, policy makers might rely on a suboptimal picture of reality in order to make their choices. And some citizens may feel the only way to make themselves heard is to demonstrate in large crowds on the street. Reality is showing us that even if the mainstream public sentiment today is to cooperate with and endorse public policies, tomorrow this consensus may become very fragile, if misunderstandings or unrest in certain parts of the population are disregarded. Monitoring public perception will allow authorities to communicate and devise interventions in ways that strengthen the positive potential of the current public health crisis while reducing any collateral damage to society.

Crucially, such monitoring can and should be done in a way that is not perceived as unwanted surveillance by citizens, but as an initiative that invites their active input.

We aim to address these gaps by establishing PubliCo, an experimental online platform built on a strong participatory citizen science component that will serve two purposes: collecting real-time data on COVID-19-related public perception, and providing personalized, timely and reliable information to the public.

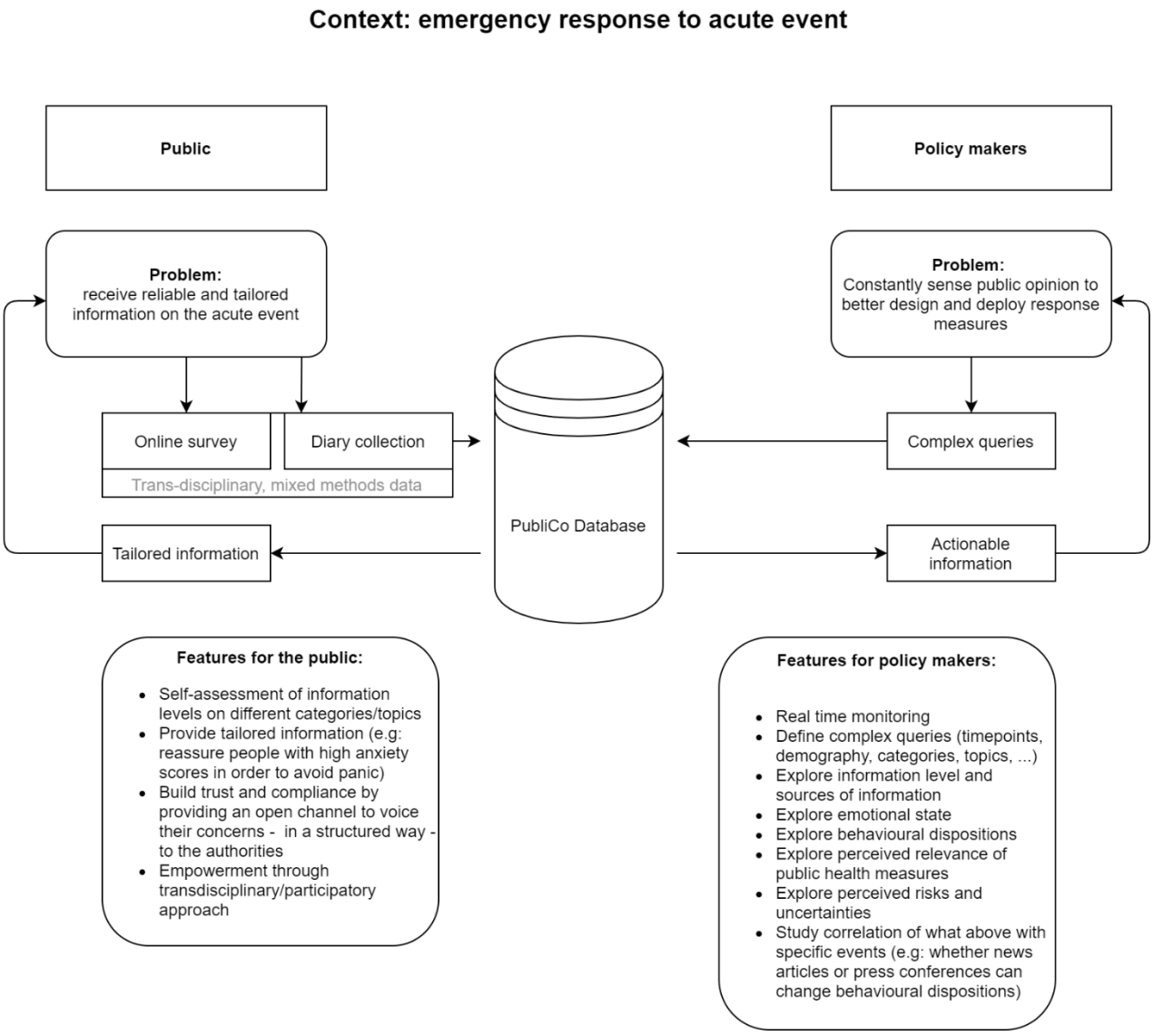

Figure 1. PubliCo conceptual structure: after completing a short survey, citizens can receive personalized information, tailored to their needs. Policy makers can study the information provided by citizens in order to conceive, deploy and evaluate more efficient mitigation and containment measures.

Methods and analysis

Concept

The project intends to facilitate well-targeted health policy-making by "feeling the pulse of the public", based on the theory that successful communication, public understanding and consent reinforce the effectiveness of public health measures [2–4]. The project combines analytical work and empirical studies using mixed methods and strong participatory elements in order to deliver a functional platform composed of three main elements: PubliCo Survey; PubliCo Diaries, PubliCo Analytics.

PubliCo Survey is the main source of quantitative information of this setup. Based on the demography and on the scores of selected subscales, citizens will rapidly obtain information that is specific to their needs. For example, people living in border regions will receive information about neighboring Countries, and people

with kids will receive information about safety measures in schools. The survey will be ongoing, providing real-time data on public perception and readiness to cooperate with public health strategies [15].

PubliCo Diaries is the main source of qualitative information in this setup. Surveys are very efficient tools to measure and analyze known or forecasted phenomena, but they lack the structural capacity for studying emerging, developing or unexpected phenomena. Moreover, diaries can provide "unique insights into the life-worlds inhabited by individuals; their experiences, actions, behaviors, and emotions and how these are played out across time and space" [16]. Users will be invited to register as citizen scientists and keep a short weekly diary in which to record their reflections on how COVID-19 and related policy measures affect their daily routine, social practices, values and priorities. In order to include also less tech-savvy segments of the population, citizen scientists can also keep their diaries offline and have the text entered by project staff afterwards. While the survey will provide information about magnitude, these texts will provide information about meaning, plus new insights on emerging, unforeseen issues. Finally, the analysis of the data gathered from the diary component will also inform the revision or generation of new survey items.

PubliCo Analytics will be the "access door" to the data collected through the survey and the diaries. It will provide information to be used for analyses directed to policy-makers, regarding information levels, behavioral dispositions, emotional states, moral preferences, and eventual correlations; for example, whether the preferences for vaccine prioritization of the elderly and of young people are the same, or whether people who have been directly on indirectly hit by the pandemic are more likely to respect preventive measures. Finally, PubliCo Analytics will contain thematically focused policy briefs, in which we will contextualize the data, interpret core findings, and suggest recommendations.

Development

The development of PubliCo will proceed in three main work packages: definition of the survey and of the information to provide through it; realization and testing of the platform; definition of the information to display through the Analytics component.

What are people looking for?

In order to define what to ask and what kind of information to provide we examined three categories of sources: Google Trends searches, publications in the general media (grey literature), and existing surveys. The aim was to capture what kind of information people are looking for, what kind of information is available in the media, and what is the focus of COVID-19 related BSSR research.

Through Google Trends we retrieved all the queries regarding COVID-19 performed in Switzerland between January 2020 and July 2020. Table 1 reports the structure of the query as suggested by Mavragani et al. [17], Google Trends data are publicly available for replication or confirmation purposes.

Query	Coronavirus + covid + 2019-nCoV + SARS-CoV2
Query type	keyword
Timeframe	01/01/20 - 27/07/20
Date of search	27/07/2020
Data source	Web searches
Location	CH (by canton)
Query category	all

Table 1. Structure of the Google trends search on COVID-19

We analysed the normalized hits per Canton over time and the top searches associated to the keywords in our list, defined by Google trends as "terms that are most frequently searched with the term you entered in the same search session, within the chosen category, country, or region" [18].

The normalized hits per week allow to see a very general national trend in the need for information about a new phenomenon: the novelty of the outbreak explains the initial spike in February/March, after which the searches normalize around a new baseline. It is important to keep in mind, as pointed out by several studies on infoveillance, that Google Trends does not provide raw numbers, but only normalized hits. The normalization of data "indicates that the values vary from 0 to 100. The value 0 does not necessarily indicate no searches, but rather indicates very low search volumes that are not included in the results." [19].

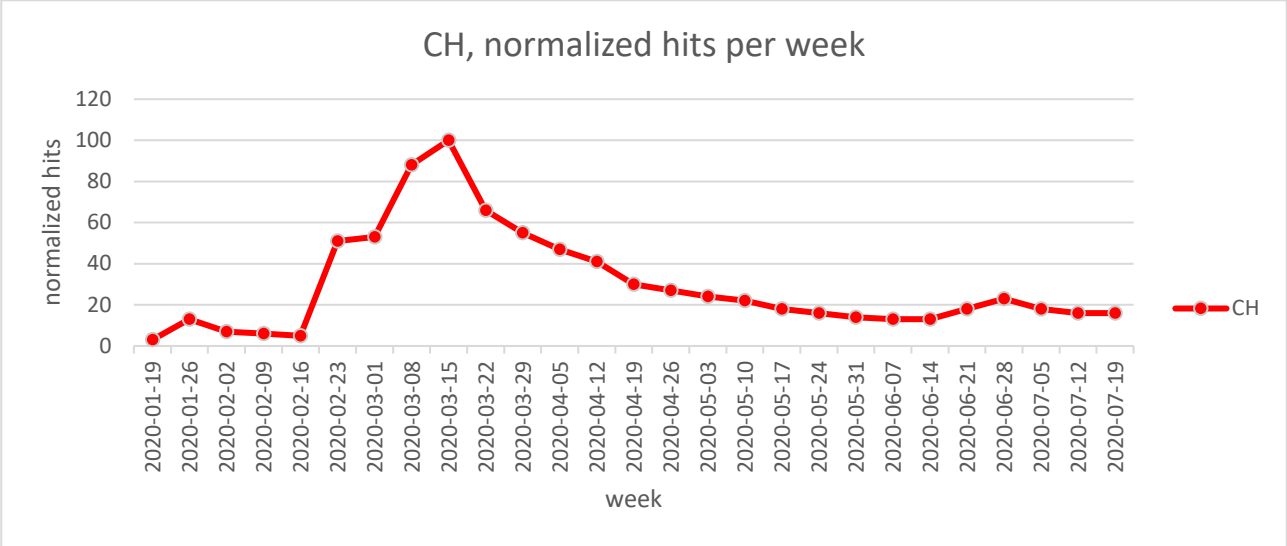

Figure 2. Normalized hits per week for COVID-19 related keywords in Switzerland.

It is worth noting the temporal distribution of the normalized hits per week: after the initial peak (mid-February to mid-April) the count seems to stabilize. A working hypothesis for interpreting this trend in information consumption is that people needed a lot of "ontological" information about a new phenomenon when it appeared (e.g: "what is coronavirus?", "what are the symptoms?"); it is reasonable to suppose that in the post-peak phase the information need steered towards live updates, new regulations and provisions, travel limitations and so on ("management information").

Top associated searches are more interesting, as they allow to see what users were looking for when looking for terms comprised in the search strategy. We hand-coded the top associated searches, assigning to each one of them a category; categories were defined bottom-up, during the coding process. Table 2 lists the categories we defined, the explanation, and some examples of searches belonging to that category.

Category name	**Explanation**	**Example**
geographical reference (place of residency)	Country, region or city the searcher lives in	"Aargau coronavirus"
geographical reference (other place)	Country, region or city that differs from the origin of the query	"Deutschland coronavirus"
official body	World health Organization, Federal Office for Public Health	"coronavirus ofsp" "who coronavirus"
quantitative information	Number of cases, number of deceased, other statistics	"cas coronavirus suisse"

news	News, either specific to a place or general.	"coronavirus news schweiz"
medical information	Information related to the diagnosis, treatment or outcome from a clinical point of view	"coronavirus symptome"
tips	Suggestions and advice	"coronavirus tipps"
live update	Live updates on the pandemic situation	"coronavirus schweiz aktuell"
general information	General information about the virus, the disease or the pandemic	"info coronavirus"

Table 2. Categories, definitions and examples of items attributed to each category.

The count of the categories of the top associated searches allows to understand what macro-topics have been perceived as most interesting (reported in Table 3). Moreover, comparing the mean of the cantonal data (column 3) to the aggregated data for the whole Country (column 2) allows to see how some macro-topics can be considered important in some areas, but without emerging in the aggregated data. It follows that, especially in Countries characterized by geographic, cultural and linguistic diversity as Switzerland, infoveillance studies using Google trends data need to be granular, using lower level data aggregation strategies.

Categories	**CH, Google Trends aggregated data**	**CH, mean per canton**	**CH, min per canton**	**CH, max per canton**	**CH, SD per canton**
geographical reference (place of residency)	9	7,23	1	11	2,83
geographical reference (other place)	5	3,38	0	6	1,98
news	3	2,12	0	4	1,18
quantitative information	1	2,04	0	4	1,56
general information	2	1,73	0	4	1,19
official body	1	1,04	0	2	0,60
medical information	2	0,92	0	2	0,63
live update	2	0,92	0	3	0,98
tips	0	0,54	0	1	0,51

Table 3. Category count and descriptive statistics.

It is important to stress that these data are generated with an observational design, not in a controlled experiment. Many variables can influence the query composition and the information consumption patterns. Nevertheless, the differences in information consumption are already sufficient to justify experimenting a system like PubliCo, allowing the delivery of personalized information.

What's in the media?

In order to understand what information people find and consume when performing these queries, we used Factiva, a news monitoring and search engine developed and owned by Dow Jones, to gather and download all the news articles published between January and July 2020 on Covid-19 and Switzerland. Factiva allows a very granular definition of the queries, and moreover has access to full text articles published by the major media outlet of the world. The query has been defined as outlined in Table 4.

Syntax	Meaning
((coronavirus or Wuhan virus or corvid19 or corvid 19 or covid19 or covid 19 or ncov or novel coronavirus or sars) and (atleast3 coronavirus or atleast3 wuhan or atleast3 corvid* or atleast3 covid* or atleast3 ncov or atleast3 novel or atleast3 corona*))	Keywords for covid19; must appear at least 3 times in the text
and ns=(gsars or gout)	Subject is "novel coronaviruses" or "outbreaks and epidemics" and "general news"
and la=X	Language is X (DE, FR, IT, EN)
and rst=tmnb	Restrict to TMNB (major news and business publications)
and wc>300	At least 300 words
and date from 20191001 to 20200801	Date interval
and re=SWITZ	Region is Switzerland

Table 4, structure of the Factiva query

The query is not limited to articles published by Swiss media, but to articles regarding Switzerland. The reason is simple: a Swiss user googling for "Schweiz Coronavirus" or for "Coronavirus Ticino" can easily find and read articles published by foreign media outlets (namely, German, French or Italian) on that topic. If the objective is capturing and describing the information trends to which people are exposed, this approach makes much more sense than limiting the analysis to articles published by Swiss media.

Factiva's field "NS" is a descriptor for the content of the article. "gsars" is defined in Factiva's documentation as "All news on Severe Acute Respiratory Syndrome", and "gout" as "The widespread occurrence of an infectious disease affecting many people or animals in a given population at the same time"; however, the way these descriptors are assigned to articles is not specified in the documentation [20] .

Finally, the query has been restricted to major news and business publications of at least 300 words. Duplicate check is performed by Factiva. Given the large amount of articles published on COVID-19, this arbitrary restriction allows to retrieve a corpus that is both meaningful and manageable, listed in table 5. Figure 3 displays the linguistical and temporal distribution of the retrieved articles.

	Results	Duplicates	Included
EN	2030	662	1368
DE	4741	1816	2925
FR	861	222	639
IT	209	52	157

Table 5, results, duplicates and included per language

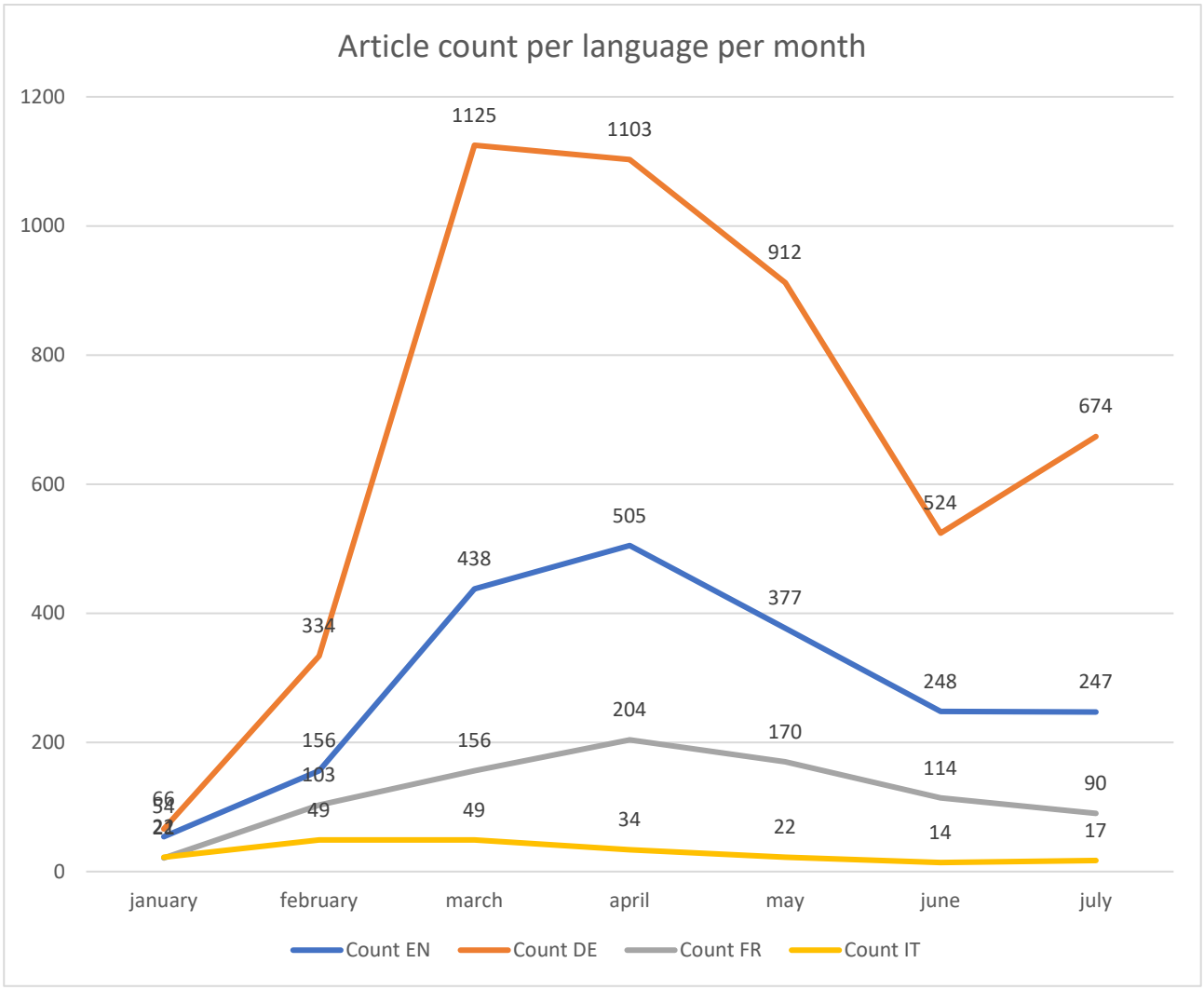

Figure 3. Article count per month. blue: English; orange: German; grey: French; yellow: Italian

We parsed the files downloaded from Factiva (unfortunately original files are not shareable due to copyright reasons) using regex rules to extract article ID, language, date, title, Author and text. Articles were then saved in a new data frame and exported as a .csv file. This approach allows more refined segmentation of the text in the natural language processing pipeline, for instance limiting the analysis to a very specific time interval. The parser is available for confirmation and replication. [21].

The NLP pipeline adopts a very simple approach, coherent with the purpose of this part of the study. The overall idea, by means of lemmatization and named entity recognition (NER), is to reduce the corpora to a list of concepts to be tracked over time and to be confronted with the results of the Google trends data analysis. We used spaCy, a widely adopted Python library that together with NLTK represents the state of the art for NLP [22]. Moreover, spaCy supports over 59 languages, making it easier to replicate this section of the study in many different linguistical contexts. The analysis software and the resulting dataset are available for confirmation and replication. [23].

The analysis of German lemmas, compared to the other linguistic subcorpora, indicates a public discourse highly focused on quantitative aspects of the pandemic (“prozent”, “million”, “milliarde”, “zahl”, “fall”).

The French subcorpus indicates a public discourse focused on describing the pandemic and its effects on people (“cas”, “crise”, “pandémie”, “personne”, “santé”). It is rather interesting to notice the opposed trends of “pandemie” and “epidemie”: the first increases by time, while the second decreases, with the swap point happening between March and April.

The Italian subcorpus appears to be focused on cases and fatalities more than the other ones ("caso", "contagiare", "contagio", "morto"). Interestingly, it is also the subcorpus in which "oms" (Italian acronym for "Organizzazione Mondiale della Sanità", i.e. WHO) ranks higher, indicating a higher attention for official WHO news and/or reports. The Italian subcorpus is also the only one in which "China" (i.e: "Cina") appears in the top 20 lemmas.

The English subcorpus, finally, seems dominated by information reported from other sources ("say"), which makes sense, given the fact that English is not an official language of the Confederation. Under the surface of lemmas not that different from those found in the other subcorpora (so, effects of the pandemic, quantitative information, information from official bodies) it is possible to notice many lemmas like "company", "group", "market", suggesting a higher attention to the economic and financial impact of the pandemic.

It appears that all the subcorpora provide, at varying levels, the following macro-categories of information:

- georeferenced information (information specific to countries, Cantons or cities);
- general information about the pandemic and about the virus;
- reports from authorities and official bodies;
- quantitative information.

Named Entity Recognition (NER) is another, more refined technique employed in text mining to conflate texts. It is widely employed in information extraction and retrieval, automatic summarization, automatic question answering and similar tasks [24–26]. NER can recognize the category (e.g. person, location, organization) of a given word, allowing to define subsets of concepts in the corpus, making it possible to understand, for example, what person or what organization is mentioned more often. In the context of this project, information extraction by means of NER serves the purpose of validating what emerged with the analysis of lemmas, allowing in the meantime to gather more fine-grained information.

The most frequent category of named entities in the German subcorpus is LOC, i.e. geographical places. This information allows to contextualize and expand what emerged in the lemma analysis: German media reported lots of quantitative information referred to "Schweiz", "Deutschland", "China", "Italien", "Frankreich", "Österreich", "USA" and so on. Importantly, the most named countries are either the neighboring ones or the ones in which the pandemic it more strongly. The only cities that appear in the top 20 entities are Zurich and Bern. "BAG", the acronym of "Bundesamt für Gesundheit" (i.e: Federal Office for Public Health) is the first organization mentioned, signifying a special attention to reports, news and deliberation of the Swiss federal body in charge of public health.

The analysis of the French named entities represents a similar situation, with lots of mentions of Countries that are either neighboring or hit severely by the pandemic ("Suisse", "Chine", "Etats-Unis", "Europe", "Italie", "France"). The first organization mentioned is "OMS" (French acronym for "Organisation mondiale de la santé", i.e. WHO), followed by "Conseil fédéral" (i.e. Federal Council) and by "OFSP" (French acronym for "Office fédéral de la santé publique", i.e: Federal Office for Public Health). The only cities that appear in the top 20 entities are Genève and Wuhan.

Also the Italian named entities go in the same direction: the most mentioned Countries are "Cina", "Italia", "Usa", "Europa", "Germania", "Svizzera", "Stati Uniti", "Spagna", "Iran", "Francia" and "Corea del Sud". "OMS" (Italian acronym for "Organizzazione Mondiale della Sanità", i.e. WHO) is the second most frequent entity, after "coronavirus". The only cities that appear in the top 20 entities are "Pechino" and Wuhan.

The English named entities follow an almost identical pattern: nearby Countries or Countries with significant outbreaks ("China", "US", "Italy", "Europe", "Germany", "UK"), the WHO, Geneva.

What are researchers looking for?

On the 14th of May 2020 the NIH Office of Behavioral and Social Sciences released a document listing "data collection instruments, including surveys, for assessing COVID-19-relevant Behavioral and Social Science (BSSR) domains for clinical or population research" [27].

In order to understand the foci of current BSS research related to COVID-19 we screened all the items of every survey in the NIH collection, classifying the topic of each one of the subscales and then grouping them in categories. We identified 6 main categories and 35 subcategories, listed in Table 6.

Category	**Topics**
Financial impact	Impacts on work & childcare; Deprivation
Social practices	Social connections; Social distancing; Social capital
Behavioral dispositions	Recent risky/protective behavior behavior; Cleaning behavior; Work behavior; Coping behavior; Interpersonal conflict; Comparison with others; Anticipated vaccination behavior; Healthful behavior; Sleep; General disruptions
Moral preferences	Willingness to distance; Federal government response; State government response; City government response
Emotional state	Depression Screening; Anxiety Screening; Stress Scale; Resilience; Emotional Regulation; Loneliness; General Emotional Impact; Worries; Obsession with Covid; General Well-being; Cognitive Well-being; Sleep
Cognitive understanding	COVID-19 symptoms; What to do if symptomatic; COVID-19 Transmission; Self-protection

Table 6. Categories and topics on which current COVID-19 BSSR research is focused.

Summarizing, through the analysis of Google Trends data and of the Factiva data we identified 5 categories of information to collect and to provide through PubliCo: demographics, cognitive understanding, behavioral dispositions, emotional state, moral orientations. The NIH BSSR collection then provided a comprehensive overview of what's going on in the field, and more importantly a solid basis to start with for defining our own survey tool.

These five categories allow us to provide users with information which is very personalized, taking into account many factors at the same time; on the policy makers side, i.e. through PubliCo analytics, we can answer complex questions like "are people who know someone who got infected by covid more likely to get vaccinated, and if so, how would they prefer the vaccine to be distributed?". Moreover, such queries can be segmented, yielding actionable information also on specific subgroups (e.g. age, residency, level of instruction).

Open science by design

We believe that, in the current context, adopting a democratic, bottom-up approach in the design and development of PubliCo would greatly improve the perception of the project, allowing also to tackle urgent and unforeseen issues [28].

Because of this assumption, every component of PubliCo is or will be publicly available: the research project, the intermediate datasets and the software with which they have been compiled, the source code, the raw data and the interpretative briefs. The only data that will be subject to manual check before release is the raw text of the diaries: collateral information such as the name of a relative, the date of a specific event or geographical references might lead to re-identification of contributors, which must be avoided.

Citizen scientists will also be involved in the development of the survey and of the information we intend to provide. This will be accomplished through the web-based project builder of the Citizen Science Center Zurich. [29].

We believe that this setup will increase trust in the project, will encourage secondary use of PubliCo data, and after a successful pilot phase will foster the implementation of the tool also in other Countries.

Data collection

Data collection will start with a pilot phase (December 2020 to January 2021), during which we will collect analytics on how the platform and its different tools are used. For this purpose, we will use a shorter version of the PubliCo survey (which now is still under evaluation through Citizen Science Center Zurich). The tool is already able to collect and provide information as by specifics, but we want to collect more bottom-up input before a final deploy.

Data collection for the diary component will already start during the pilot phase. Participants will be given a brief guide to the diary method, which will inform them about the openness of the method (e.g., without concerns about spelling and grammar). The guide will ask them to jot down a) their experiences and thoughts from the beginning of the pandemic to the current day and b) their everyday worries, emotions, risks, experiences, decisions and actions during and/or after the pandemic in at minimum a weekly rhythm for a duration of at least 4 weeks. This will allow us "to document changes in values, attitudes, knowledge and behavior" [30].

In order to increase the user base PubliCo will be disseminated through a series of press releases following the kick-off event (December 2020); moreover, the outboarding section will invite the users to share the tool further via social media, emails or similar systems, capitalizing on snowballing. To increase the dissemination even further, official channels like the automatic SMS sender of the Federal Office of Public Health are another good option we will investigate.

Data collection will be iterative and will proceed for at least two years. We expect the tool to be refined and enhanced as data collection and analysis moves forward. The current version of the survey is available at www.publico.community.

Data analysis

Results from the online survey will be analyzed in multiple ways. Users will have direct feedback for certain variables (e.g: information level, behavioral dispositions), including scores and correct information based on responses to knowledge questions but also basic descriptive statistics (means and frequencies) for all users and specific sub-groups or respondents from specific cantons.

In addition, PubliCo Analytics will allow to investigate complex and flexible queries, including factors like time of the answer (including trends), geographical information, demographic information, and subscale scores.

Project researchers will also analyze results for periodic policy briefs. Questions to be examined will vary over time and will include basic descriptive statistics for the different domains included in the survey (knowledge, emotional state, behavioral dispositions and moral preferences), sub-group analyses by geographical area and target groups, and correlation analyses. Questions to be examined through correlation analysis include:

- What is the relationship between participant knowledge and willingness to comply with public health restrictions?
- What is the relationship between participant knowledge and emotional state?
- What is the relationship between participant's emotional state and their willingness to comply with public health restrictions?

- What factors influence participants' moral preferences?

These and other questions will be analyzed using regression analysis with a significance level of $\alpha = 0.05$.

The diary narratives will be anonymized and analyzed in conjunction with the ongoing data collection by means of thematic analysis [31], using the Software MAXQDA [32].

Selected data will be displayed in PubliCo Analytics in a visually appealing form (e.g. infographs, live maps). Advanced analytics will be employed whenever possible (NLP for text elements, predictive modelling of, e.g., public behavior in case of new measures taken). Data collection will be adapted to how the situation evolves, taking up emerging themes (e.g. a new vaccine not yet available to all; balancing work requirements and protection of persons with risk factors). Core findings alongside recommendations will be published in thematically focused policy briefs in which we will contextualize the data, interpret core findings, and suggest recommendations.

Ethics and dissemination

The ambitions of PubliCo are high. The first aim is to deliver personalized information in the context of public health emergencies, which is widely recognized as a crucial activity. Providing personalized information can be potentially problematic: In the context of subscales like "sources of information" it can be rather straightforward, simply pointing the user to wrong answers in knowledge quizzes and to reliable sources like the WHO or national/European official information outlets [33]. Some uneasiness remains, however, around making assumptions about citizen's informational needs and possibly contributing to knowledge "bubbles". Providing personalized information from subscales regarding emotional response, moral preferences or mental wellbeing can be even more challenging. For these topics we will rather provide a comparison between individual scores and national means. In this sense, it is fundamental to clarify the descriptive nature of the scores without any claims as to what the norm should be (is-ought problem). However, the final strategy needs to be defined with expert advisors and citizen scientists after evaluating potential outcomes.

Our second aim is to provide real-time analyses of the current situation much richer and more nuanced – and thus more effective – than the sheer numbers of citizens tested, infected, recovered, or deceased. We aim to capture and describe the bio-psycho-social dimension of living in a public health crisis (as perceived by citizens), while providing useful information to both citizens and policy makers. Many passages, from the analysis of written diaries to the automated analysis of selected subscales, will be automatized by means of NLP and other related AI applications. These techniques will ensure that the platform is more cost effective and that results of analysis and actionable information are available faster.

The Swiss Cantons are affected in different ways from the COVID-19 pandemic. Our approach, comparing geo-located data, might unveil relevant differences in behaviours and attitudes that could correlate with the course and the severity of the pandemic itself. Because of this, we will collect some demographic information (personal data, potentially also sensitive as defined in the Law on Information and Data Protection (IDG) par. 3 of the Canton of Zurich) and some information about personal philosophical or religious beliefs (sensitive data as defined in IDG par. 3).

We assessed the potential risks of the project. The most prominent category of risks is connected to re-identification of participants. Because of this, the survey component is completely anonymous by design (not even the IP address is collected) and the diary component is pseudonymous by design (we can attribute diaries to users, but we cannot attribute users to persons). The only remaining concrete risk for re-identification is posed by what users could write in the diaries. Because of this, we took extra care in

planning the access, use and management of this category of data. We are confident that the instrument is safe from a data protection point of view.

Potential event	Potential consequences	Type of harm	Severity (1-5)	Likelihood (1-5)
Re-identification of a participant	Participants can feel betrayed by the data controller and lose trust in research/society	Psychological	2	1
Re-identification of a participant	Participants with controversial opinions could lose their jobs when these are considered particularly dangerous by their employers	Economical	3	1
Re-identification of a participant	Participants with controversial opinions could be rejected and isolated from the societies of which they are part	Social	3	1
Re-identification of a participant	Participants with controversial opinions could be physically assaulted because of their opinions	Physical	5	1
Morally problematic questions	Participants can be upset when asked about morally problematic topics (e.g. allocation of scarce resources) especially if directly touched by the issue at stake	Psychological	2	3

Table 7. Risk assessment of PubliCo.

The potential harms generated by the project fall in two categories: re-identification (and thus attribution of specific opinions to specific persons) and morally problematic questions.

In order to mitigate the first category of risk, data will be treated with utmost discretion: all the data will be stored in virtual machines hosted in the UZH's data center, with access restricted to the project members. The chances of identification, in the eventuality of a data leak, are very low. As stated above, personal sensitive information resulting from the diaries will be pseudonymized. The identification keys will be encrypted with SHA-256 functions and archived for at least ten years on the servers of the IBME. It is important to point out that that kind of encryption is military-grade and has never been broken.

In order to mitigate the second category of risk we will discuss the whole survey tool with expert advisors, but also with citizen scientists, in order to get double feedback on the issues involved. That said, the impact would still be low, and more importantly the distressed user can interrupt or end participation at any time.

The very nature of this project implies another general risk: in a totalitarian context the tool we are developing could be used to spy on people. This is a potential risk we cannot mitigate for other countries. For Switzerland, the whole infrastructure of the project is in fact built keeping in mind a transparent and democratic approach, important in general in the scientific enterprise, but fundamental in a context in which the data yielded from the system are used in order to make decisions impacting the public.

Overall, participants do not have an immediate personal benefit beyond the insights gained through the survey experience and feedback, but do have a long-term community benefit resulting from the tool being used to deploy public health measures that consider and take into account (even critically) their preferences. Therefore, we consider the risk-benefit balance justifiable.

Limitations

This design has two main limitations. We are aware that our approach is very much focused on public perception rather than on observational data of real practices and we acknowledge a limitation that there may be discrepancies between opinions, attitudes and behavioral dispositions and what people actually do. On the other hand, we think much insight is to be gained already from what people are in principle agreeable to or what they will consider inacceptable.

The second limitation regards the information to provide at the end of the survey: for some topics, e.g. the concrete risk posed by COVID-19, it is difficult to find solid figures and the way they are communicated can generate problems and misunderstandings. In this sense we have opted for a different approach: users with high scores in the information level topics will be pointed first to the official information provided by the Federal Office of Public Health, and secondly to PubMed queries designed to yield systematic reviews or meta-analyses on the specific topic. This way, following once again an open science spirit, citizens will be able to access the literature (or at least the abstract, for non-open access articles) and confront themselves with the complexity of the topic. Which, in some situations, is a fundamental insight to communicate.

Acknowledgements

the Authors would like to acknowledge Samuel Giacomelli, Luca Fedrizzi, Claudio Postinghel, Luca d'Incà, Giulio Michelon, and the whole Belka team for their fundamental contribution to the development of the project's software.

Funding

This project is supported by the Swiss National Science Foundation, project number 31CA30_195905, by the WHO, and by the Swiss State Secretariat for Education, Research and Innovation (SERI)

References

1 swissinfo.ch. 1,000 people gather for unofficial football match. swissinfo.ch. 2020.https://www.swissinfo.ch/eng/corona-restrictions-broken-_1-000-people-gather-for-unofficial-football-match-/45778644 (accessed 27 Nov 2020).

2 Lee ST, Basnyat I. From press release to news: mapping the framing of the 2009 H1N1 A influenza pandemic. *Health Commun* 2013;**28**:119–32. doi:10.1080/10410236.2012.658550

3 Sandell T, Sebar B, Harris N. Framing risk: communication messages in the Australian and Swedish print media surrounding the 2009 H1N1 pandemic. *Scand J Public Health* 2013;**41**:860–5. doi:10.1177/1403494813498158

4 Lin L, Savoia E, Agboola F, *et al.* What have we learned about communication inequalities during the H1N1 pandemic: a systematic review of the literature. *BMC Public Health* 2014;**14**:484. doi:10.1186/1471-2458-14-484

5 van der Weerd W, Timmermans DR, Beaujean DJ, *et al.* Monitoring the level of government trust, risk perception and intention of the general public to adopt protective measures during the influenza A (H1N1) pandemic in The Netherlands. *BMC Public Health* 2011;**11**:575. doi:10.1186/1471-2458-11-575

6 WHO. Novel Coronavirus(2019-nCoV) Situation Report - 13. 2020.https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf (accessed 4 May 2020).

7 WHO. Coronavirus disease (COVID-19) advice for the public: Myth busters. 2020.https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters (accessed 4 May 2020).

8 Plough A, Bristow B, Fielding J, *et al.* Pandemics and health equity: lessons learned from the H1N1 response in Los Angeles County. *J Public Health Manag Pract JPHMP* 2011;**17**:20–7. doi:10.1097/PHH.0b013e3181ff2ad7

9 Kaplan G, Baron-Epel O. Personal needs versus national needs: public attitudes regarding health care priorities at the personal and national levels. *Isr J Health Policy Res* 2015;**4**. doi:10.1186/s13584-015-0010-2

10 SteelFisher GK, Blendon RJ, Ward JRM, *et al.* Public response to the 2009 influenza A H1N1 pandemic: a polling study in five countries. *Lancet Infect Dis* 2012;**12**:845–50. doi:10.1016/S1473-3099(12)70206-2

11 SRF. Umfrage der SRG - So denkt die Schweiz über die Corona-Krise. Schweiz. Radio Fernseh. SRF. 2020.https://www.srf.ch/news/schweiz/umfrage-der-srg-so-denkt-die-schweiz-ueber-die-corona-krise (accessed 7 Jun 2020).

12 tagesschau.de. DeutschlandTrend extra: Deutsche finden Versammlungsverbot gut. tagesschau.de. 2020.https://www.tagesschau.de/inland/deutschlandtrend-extra-blitzumfrage-103.html (accessed 7 Jun 2020).

13 Betsch C, Wieler L, Bosnjak M, *et al.* Germany COVID-19 Snapshot MOnitoring (COSMO Germany): Monitoring knowledge, risk perceptions, preventive behaviours, and public trust in the current coronavirus outbreak in Germany. Published Online First: 3 March 2020. doi:http://dx.doi.org/10.23668/psycharchives.2776

14 Gould DJ, Drey NS, Millar M, *et al.* Patients and the public: knowledge, sources of information and perceptions about healthcare-associated infection. *J Hosp Infect* 2009;**72**:1–8. doi:10.1016/j.jhin.2009.01.024

15 Wäscher S, Salloch S, Ritter P, *et al.* Methodological Reflections on the Contribution of Qualitative Research to the Evaluation of Clinical Ethics Support Services. *Bioethics* 2017;**31**:237–45. doi:10.1111/bioe.12347

16 Milligan C, Bartlett R. Solicited Diary Methods. In: Liamputtong P, ed. *Handbook of Research Methods in Health Social Sciences*. Singapore: : Springer 2019. 1447–64. doi:10.1007/978-981-10-5251-4_15

17 Mavragani A, Ochoa G, Tsagarakis KP. Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. *J Med Internet Res* 2018;**20**:e270. doi:10.2196/jmir.9366

18 Google. FAQ about Google Trends data - Trends Help. 2020.https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052 (accessed 27 Nov 2020).

19 Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. *JMIR Public Health Surveill* 2019;**5**:e13439. doi:10.2196/13439

20 Dow Jones Factiva. Dow Jones Developer Platform - COVID-19. 2020.https://developer.dowjones.com/site/docs/building_queries/data_selection_samples/covid_19/index.gsp (accessed 27 Nov 2020).

21 Spitale G, Merten S, Biller-Andorno N. *Factiva parser and NLP pipeline for news articles related to COVID-19*. Zenodo 2020. doi:10.5281/zenodo.3991821

22 spaCy. Facts & Figures. SpaCy Usage Doc. 2020.https://spacy.io/usage/facts-figures (accessed 27 Nov 2020).

23 Spitale G, Merten S, Biller-Andorno N. Lemmas and Named Entities analysis in major media outlets regarding Switzerland and Covid-19. 2020. doi:10.5281/zenodo.4036071

24 Gupta V, Lehal GS. A Survey of Text Mining Techniques and Applications. *J Emerg Technol Web Intell* 2009;**1**. doi:10.4304/jetwi.1.1.60-76

25 Kaur D, Gupta V. A survey of Named Entity Recognition in English and other Indian Languages. *Int J Comput Sci Issues* 2010;**7**:239–45.

26 Nadeau D, Sekine S. A survey of named entity recognition and classification. *Lingvisticæ Investig* 2007;**30**:3–26. doi:10.1075/li.30.1.03nad

27 NIH Office of Behavioral and Social Sciences Research. COVID-19 OBSSR Research Tools. 2020.https://www.nlm.nih.gov/dr2/COVID-19_BSSR_Research_Tools.pdf (accessed 10 Jun 2020).

28 Smart P, Holmes S, Lettice F, *et al.* Open Science and Open Innovation in a socio-political context: knowledge production for societal impact in an age of post-truth populism. *RD Manag* 2019;**49**:279–97. doi:https://doi.org/10.1111/radm.12377

29 Citizen Science Center Zurich. CS Project Builder - PubliCo. Citiz. Sci. Cent. Zurich. 2020.https://lab.citizenscience.ch/en/project/231 (accessed 27 Nov 2020).

30 Constant N, Roberts L. Narratives as a mode of research evaluation in citizen science: understanding broader science communication impacts. *J Sci Commun* 2017;**16**:A03. doi:10.22323/2.16040203

31 Hyers L. *Diary Methods: Understanding Qualitative Research*. Oxford University Press 2018. https://www.oxfordscholarship.com/view/10.1093/oso/9780190256692.001.0001/oso-9780190256692 (accessed 7 Jun 2020).

32 Woolf NH, Silver C. *Qualitative Analysis Using MAXQDA: The Five-Level QDATM Method*. Routledge 2017.

33 EU vs Disinformation. EU vs DISINFORMATION. EU Vs DISINFORMATION. 2020.https://euvsdisinfo.eu/ (accessed 10 Jun 2020).