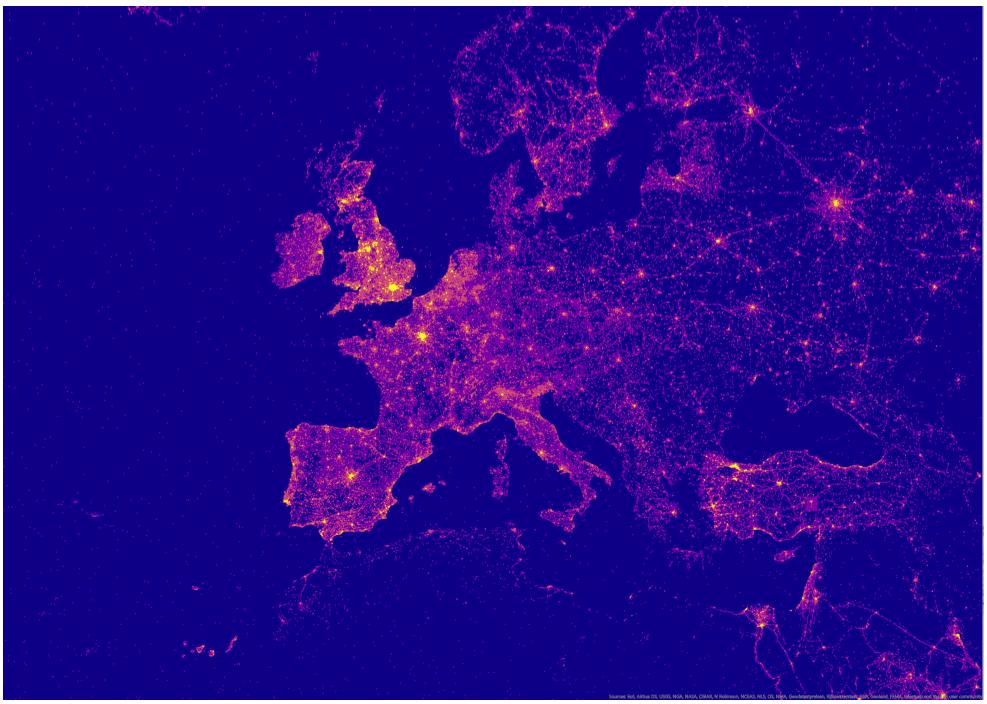
Final Report for SSPH+ project
Social Media Mental Health Surveillance – M-HEALTH

Funding period October 2018 to September 2019



Co-applicants
Oliver Gruebner, Milo Puhan, UZH EBPI
Emiliano Albanese, USI

Table of content
Final report

Attachments:
Data management plan
Data sharing and confidentiality agreement

Figure caption
The geography of Twitter tweets in January 2015

Final Report

This report describes the major achievements of the M-HEALTH project, funded by SSPH+ over the time frame between October 2018 to September 2019. The M-HEALTH project's goals, challenges, and deliveries are outlined in detail below. Some challenges had to be overcome, which significantly delayed parts of the tasks. However, the major goals as originally defined were achieved.

Personnel

The M-HEALTH project brought together scholars from digital health geography / spatial epidemiology (Gruebner, co-lead, UZH, Zurich), psychiatry (Albanese, co-lead, USI, Lugano), epidemiology (Puhan, UZH), and data science (Sykora, Loughborough University, Leicestershire, UK).

Furthermore, Ivor Mardesic (geographic information science) started to work on the project goals on a 20% assistant position located at UZH from March 2019 and Marta Fadda (psychology) started on a 10% postdoc position located at USI from May 2019. In addition, Suzanne Elayan (information science, Loughborough University, UK) and Tamar Edry (geographic information science, UZH) joined the team in June and September 2019, respectively and were financed by additional University funds.

Joint activities

The team had regular Skype meetings and onsite meetings in Zurich and Lugano during which we were able to discuss all procedures and challenges to success. The list below characterises the overall structure and nature of the meetings within the M-HEALTH project, although it was not possible to always maintain the schedule for each and every week due to availabilities and conference travels of individual team members,

- Weekly **on-site meetings** at UZH (Ivor Mardesic, Tamar Edry, Oliver Gruebner)
- Weekly **on-site meetings** at USI (Marta Fadda, Emiliano Albanese)
- Bi-weekly **Skype meetings** (Marta Fadda, Ivor Mardesic, Martin Sykora, Suzanne Elayan, Oliver Gruebner), with sporadic participation by Emiliano Albanese according to needs
- **On-site joint meetings** at UZH, with the UZH Department of Data Protection and with the data protection officer of canton of Zurich (Marta Fadda, Oliver Gruebner, Ivor Mardesic).
- **Summer school on digital spatial epidemiology** at UZH, August 5-9 (with Suzanne Elayan, Marta Fadda, Oliver Gruebner, Ivor Mardesic, and Martin Sykora).
- **Project progress and advisory board meeting** on August 9th 2019 (with Emiliano Albanese, Suzanne Elayan, Marta Fadda, Oliver Gruebner, Ivor Mardesic, Martin Sykora, and Milo Puhan, External guests: Sven Lautenbach, Sara Fabrikant).
- **Talk introducing the M-HEALTH project** and related studies at the SSPH+ summer school in Lugano on August 27th 2019 (Oliver Gruebner).

Project outcomes

Within the M-HEALTH project, we applied three work packages (WP):

WP 1) Administering geo-social media data over five years.

WP 2) Enriching the original tweet data with sentiment analysis.

WP 3) Setting up a repository with the enriched geo-social media data and providing geographic information on a website (details below).

Administering geo-social media data over five years. We 1.1) obtained geo-referenced Twitter data for Europe for the years 2014-2018 available from the Harvard Centre for Geographic Analysis (CGA). These data included the IDs for each tweet captured by CGA in the given area and time frame. CGA also provided us with a script for downloading the full tweets from the official Twitter API, a process that is called rehydration. The rehydration was done at the Loughborough University and started in December 2018. The rehydration process involves sending requests to the official Twitter API /statuses/lookup endpoint[1] with tweet IDs, which are 18-19 long numeric IDs

[1] See documentation – https://developer.twitter.com/en/docs/tweets/post-and-engage/api-reference/get-statuses-lookup.html

(e.g., 1037658048414842882), uniquely identifying a tweet, where Twitter responds with the tweet and related metadata, when available. On average, just under 30% of all tweet IDs did not receive a Twitter response[2], which is a commonly encountered attrition rate when rehydrating tweets, and is due to any of the following reasons: (i) tweets may have been deleted by the user or removed by Twitter, (ii) Twitter users have switched their account from public to private, or (iii) an account has been suspended[3]. Over the years, Twitter changed their API tweet and user object response data structures, with some new fields returned, which also needed to be incorporated into the rehydration process/script by the M-HEALTH team.

While rehydrating the tweets, several challenges delayed the process; e.g. rate restrictions imposed by Twitter or errors found in the Twitter API interface. Due to these, the rehydration script had to be altered, which prolonged the overall download time. By the time of writing this document (October 2019), the 2014 dataset had yet to be fully rehydrated for these reasons, with completion expected between December 2019 and February 2020, subject to Twitter API download rates. **Once the 2014 data is rehydrated, the M-HEALTH team will add it to the repository**. Due to Twitter's geo-referencing settings and policies (geo-reference was by default on in 2014 and there was less awareness about it), the 2014 data will be the largest, and could possibly equal in size the combined 2015-2018 data. For the years 2015, 2016, 2017, and 2018; our data contained 192,927,972 unique tweets that were produced by 4,077,843 unique Twitter user IDs that were downloaded at a rate of approximately one million per day, considering API and/or rehydration server down-time, as well as immediate post-processing. When interpreting the numbers of tweets and Twitter user IDs, it should be kept in mind that it is nearly impossible to ascertain how many real persons this accounts for. Multiple persons might use a single user id/Twitter account, a single person might use multiple user id's (i.e., Twitter accounts), or there may be bots in the data (i.e., accounts that operate according to a computer program).

Furthermore, 1.2) we pre- and post-processed the rehydrated tweet data, including inspecting the data for known bots (Twitter accounts operated by a computer program) and other non-personal tweets (commercial information, etc.) and marked the tweets respectively. We ensured internal consistency through removing duplicates and any instances where the latitude or longitude of a tweet was missing. On average, between 0.5% and 1% of rehydrated tweets were removed during pre-processing. A simple screening for bots revealed that within our data there might be, on average, between 10% and 40% bot tweets in the total tweet population. For example, such bots are automated weather stations, third party applications that compile tweets (such as a geo-referenced check-in with Foursquare), or more malicious examples that flood the Twitter space essentially with noise. Other approaches included looking up popular weather station or check-in software and screening for its use in the "source" column of the Twitter API (e.g. Sandaysoft Cumulus for automated weather stations). We also noted "check-in" tweets based on their tendency to follow a generic wording for each tweet (e.g. for Foursquare: Always starts with "I am at…", and ends with a URL.) and labelled them accordingly. In a subset for tweets, geolocated in Switzerland, we calculated the likelihood of a Twitter user to be a real person or a bot by applying a more computation intensive method named Botometer[4]. This method identified 371 (2%) out of 17,018 Swiss users, to be suspect bot like accounts, and were marked as such.

In addition, we 1.3) worked with the UZH Department of Data Protection and the data protection officer of the canton of Zurich between May and June 2019 to ensure conformity of these data with regard to the General Data Protection Regulation (EU-GDPR) and the Act on Information and Data Protection of the Canton Zurich (IDG). In cooperation with the UZH Department of Data Protection, we developed (1) a data sharing and confidentiality agreement for all parties involved and for future research collaborates interested in the data and (2) a data management plan (DMP) to describe the life-cycle of the data. The data sharing and confidentiality agreement is a formal contract that documents what data are being shared between the M-HEALTH project principal investigators and third parties (researchers interested in the data) and how the data can be accessed and analysed. The goal is to ensure the highest confidentiality and privacy standards in any future uses of the data and to make sure both the provider of the data and the agency receiving the data do not incur in any miscommunication. The document specifies the intended use of the data, any constraints on use of the data, and how to ensure data confidentiality and security. The DMP was developed according to the FAIR Data Principles and the SNSF (Swiss

[2] Attrition rate per time-period; 2018 – 30.81% (possibly due to Twitter's pro-active account suspension initiative, see [3] below), 2017 – 20.53%, 2016 – 26.65%, 2015 – 34.75%.
[3] Twitter has been suspending accounts actively, especially throughout 2018, as described in this official Twitter blog post https://blog.twitter.com/official/en_us/topics/company/2018/how-twitter-is-fighting-spam-and-malicious-automation.html
[4] https://botometer.iuni.iu.edu/#!/, in particular the "cap_universal" probability score was used to identify any accounts behaving like bots (using a threshold of 0.5 for the binary classification, as recommended in https://www.nature.com/articles/s41467-018-06930-7)

National Science Foundation) guidelines on the structure and content of a data management plan. The document outlines how data is generated, collected, documented, shared, stored and preserved, by anticipating the main ethics, legal and security issues that may emerge during the data life-cycle and providing detailed actions on how to deal with them. The DMP was approved by the UZH Department of Data Protection. We also applied for ethical clearance with the Ethics Committee of the Faculty of Arts and Social Sciences of UZH.

Original tweet data enriched with sentiment analysis. As part of WP 2, we "enriched" the data with basic emotion detection. In 2.1), we applied EMOTIVE[5] on the fly during rehydration and estimated for each downloaded English language tweet, whether it contained one of the eight basic emotions: anger, confusion, disgust, fear, happiness, sadness, surprise, and shame and labelled it accordingly. However, the emotion detection we used was applicable only to English language tweets so we also translated a Swiss (German language) subset to gain more insight. For this, we 2.2) used neural machine translation to translate German language tweets found in Switzerland into English language, and subsequently applied EMOTIVE on them also. EMOTIVE basic emotion detection rates for original English language tweets varied between 4 and 12% in the dataset, while for the translated Swiss German tweets it varied between 2% and 4%. In addition, we 2.3) checked samples of the tweets for internal consistency.

Repository with the enriched geo-social media data and geographic information on a website. We 3.1) set up a fully secured geo-data repository on an encrypted and password protected hard drive at UZH for the final data set including tweets with geolocation information, time, tweet texts, language as identified by Twitter, translations of the texts where applicable, basic emotions for English language and translated tweets, along with other information available in Twitter and data derived during dataset pre-processing. The final geo-data repository is maintained on localized computer systems of the UZH. It is fully encrypted with state-of-art AES 256-bit standard encryption and access to the repository is limited to one member of the M-HEALTH team at UZH. The repository itself has been established using the open-source programming language "R" and follows a directory-based database schema. Upon pre-processing of the rehydrated Twitter data, each tweet is assigned a new unique, primary key. Based on this primary key, the repository enables sub setting of the data for further use and dissemination while maintaining reproducibility and consistency of the original dataset. On a basic sub-division, the repository is divided into yearly directories (one for each year 2014-2018), and further subdivided into monthly directories.

Upon completion of the rehydration process as outlined under 1.1 (the 2014 dataset is still being rehydrated), the final dataset will consist of original geo-referenced Twitter tweets for the time frame between 15.03.2014 to 06.09.2018 for the wider European Region. Each tweet contains geo-location information (geographic coordinates), a time stamp for the time the tweet has been sent, the tweet text, the identified language of the tweet text as identified by Twitter, along with other information made available by the public Twitter API.

The study region was defined by a bounding box of geographic coordinates (lower left coordinates longitude=-34.402 W, latitude=24.388 N; upper right coordinates longitude=49.935 E, latitude=72.091 N) and includes parts of the Middle East and North Africa, Central Asia, and European Russia (Figure 1).

The tweets stem from 66 countries: Albania, Andorra, Armenia, Austria, Azerbaijan, Belgium, Bulgaria, Bosnia and Herzegovina, Belarus, Switzerland, Cyprus, Czech Republic, Germany, Denmark, Algeria, Egypt, Spain, Estonia, Finland, France, United Kingdom, Georgia, Greece, Croatia, Hungary, Ireland, Iran, Iraq, Iceland, Israel, Italy, Jordan, Kazakhstan, Kosovo, Kuwait, Lebanon, Libya, Liechtenstein, Lithuania, Luxembourg, Latvia, Morocco, Monaco, Moldova, Macedonia, Mali, Malta, Montenegro, Mauritania, Netherlands, Norway, de facto and de jure Palestinian Territories, Poland, Portugal, Romania, Russia, Saudi Arabia, San Marino, Serbia, Slovakia, Slovenia, Sweden, Syria, Tunisia, Turkey, Ukraine and Vatican. However, only parts of contiguous Russia, Kazakhstan, Iran, Saudi Arabia, Algeria, Mali, and Morocco/Western Sahara are included in the study region. Hence, not all tweets from these countries are included in the repository.

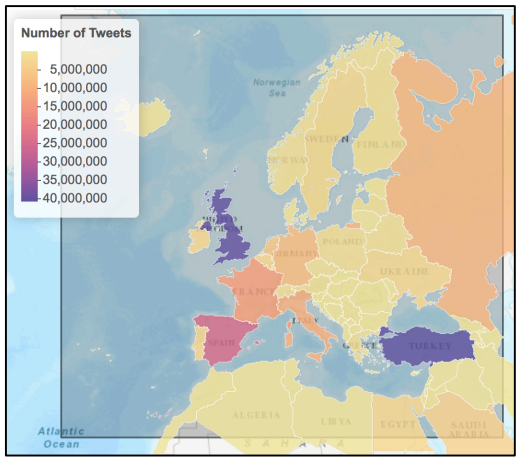


Figure 1: Bounding box of the study region.

[5] http://emotive.lboro.ac.uk

The full numbers for tweets and unique users per year in our data can be seen from Figure 2.

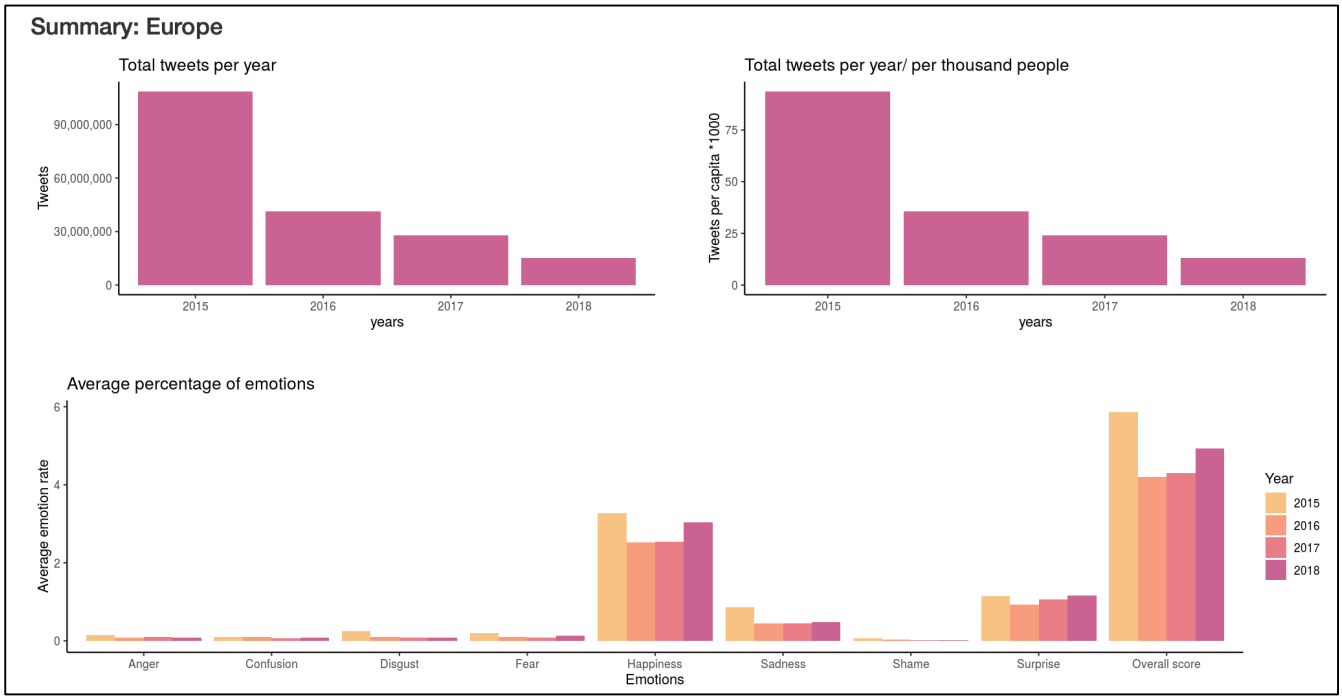


Figure 2: Data summary for the entire dataset across the wider European Region

Furthermore, we 3.2) documented all data and procedures in a data dictionary and R Markdown document for reproducibility and provided meta information attached to the processed and secured files. The meta-data is provided for each year group; including total tweets, users, unique users, temporal and spatial variance, proportion of probable bot accounts, median centroid, emotive rates, language prevalence, etc.

In addition, we 3.3) provided descriptive statistical information and exploratory spatial data analysis results in forms of tables, graphs, and maps highlighting the spatial distribution of emotions across the European Region and Switzerland. This information is available on an interactive website that also explains the project's aims along with a statement of the data sharing policies and funding (Figure 3).

The project website can be reached under: https://givauzh.shinyapps.io/tweets_app/

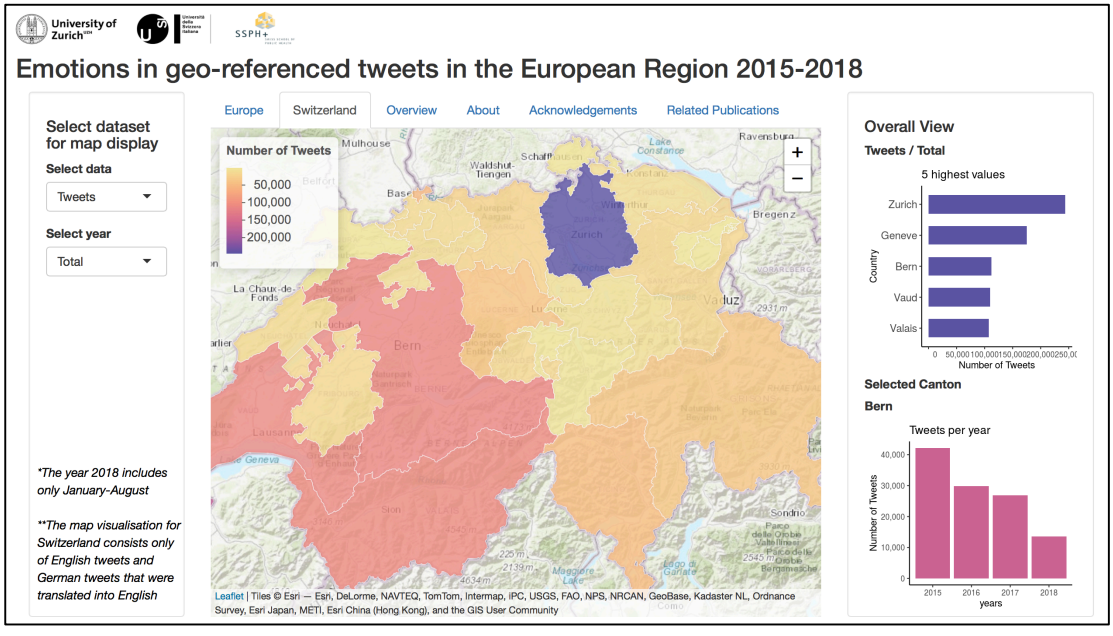


Figure 3: The project website providing information about tweets over four years. The website contains information on tweets and emotions across the wider European Region and Switzerland, along with further information on the project and its funder. It can be reached under: https://givauzh.shinyapps.io/tweets_app/

Further project outcomes and outlook

Summer School Digital Spatial Epidemiology[6]**.** This course was taking place at UZH for the first time on August 5-9th 2019 and brought together scholars from the M-HEALTH project to moderate and instruct at this event. The one-week seminar addressed spatial epidemiological approaches to social media data similar to the M-HEALTH project. We concentrated on state-of-the-art multivariable statistical and spatial statistical modelling to health outcomes as identified in geo-referenced Twitter data and associations with socio-ecological factors of urban contexts. We combined theoretical and lab work on statistical analysis and spatial-epidemiological modelling techniques in a vibrant international and interdisciplinary setting.

Contributed to UZH DSI workshop on text analysis in health[7]**.** The 2-day workshop aimed at identifying different data sources, related data analytics problems and research consequences vis-a-vis text analytics in health and took place at the UZH, September 12-13th 2019. Combining brief presentations from the participants who work with health-related digital texts, and several keynote lectures, the workshop provided a forum for exchange and expert inputs about text analytics for health. Team members Oliver Gruebner, Tamar Edry, and Ivor Mardesic contributed to this UZH DSI event by event co-organising, active participation, and by presenting the M-HEALTH project, respectively. In general, we could generate great visibility of the rich data source that were made available through the SSPH+ funded M-HEALTH project and could further attract potential collaboration partners for future research grant proposals based on this data resource to be submitted to SNSF.

Contributed to LESSON 2019 workshop[8]**.** Team member Marta Fadda presented the "Ethical and legal issues involved in conducting studies on mental health using social media data. Examples from the M-HEALTH project" at the 1st International Workshop on Legal and Ethical Issues in Crowdsourced Geographic Information (LESSON 2019) at the UZH, October 10th 2019. This presentation highlighted the main ethical challenges that emerged when planning to set up a repository with the enriched geo-referenced social media data and the strategies we intend to pursue to limit any possible risks for the users involved. The talk was enriched with information that emerged from our team's collaboration with the UZH Department of Data Protection and the data protection officer of the Canton of Zurich between May and June 2019. Our goal was to offer practical advice on how to secure ethical approval for thematically and methodologically analogue studies while contributing to the current debate on the ethical, legal, and social implications of research with georeferenced social media and big data.

Developed a new SSPH+ spring school Big Data in Public Health. We successfully applied to the SSPH+ call for a new course on big data in public health that will be held for the first time at the UZH in May 25-29, 2020. Big data approaches raise high hopes, promising that this new form of data mining and analyses will significantly improve public health research and action. The proposed course will draw substantially on the M-HEALTH project and will give a broad overview of potentials and limitations of big data from multiple public health disciplines. Definitions and concepts particularly from epidemiology, psychology, data science, digital health geography, and health policy and law are introduced and discussed in keynote lectures to evaluate the public health relevance of big data. We will also examine key legal and ethical challenges from a broader public health perspective. Subsequently, we will explore emerging methods in big data analysis and their application for advancing public health research, with a particular focus on mental health in mock sessions (i.e. case-based learning), practical exercises (i.e. working with social media data in the statistical software R), and in mini projects (i.e. group assignments extending on the lab sessions) during the course. By the end of the course, the participants are familiar with public health relevant definitions and challenges of big data approaches from major public health disciplines. They are able to address key issues in research and the broader public health discourse. Specifically, participants will: 1) Gain a broad understanding of the relevance and potential for big data to advance public health research, 2) Apply novel methods in big data analysis towards addressing important public health challenges, and 3) Develop a project with potential for real-world public health impact using big data methods.

Planned scientific manuscripts. The lessons learnt and the experiences gained especially during work package 1 of the M-HEALTH project need to be documented and shared with the research community. In particular, we aim at describing the process by which we identified the main ethical and legal issues related to this project and the strategies we developed to deal with them, especially in relation to establishing the geo-repository, publishing related data on a website, and data sharing opportunities for future collaborations with scientific scholars. The article will identify the challenges of doing research with big social media data and, in particular, location-specific

[6] http://spatialepidemiology.mystrikingly.com
[7] https://www.dsi.uzh.ch/en/research/challareas/health.html
[8] https://www.cs.nuim.ie/~pmooney/lesson2019/

data, by providing the concrete example of working with Twitter data in the Swiss legal context and by describing the counter-actions we undertook. Considering the specific features of our data, this will be the first contribution of a successful and well-detailed plan to deal with the ethical and legal issues of working with geo-referenced social media data. We aim to submit the manuscript to a peer-reviewed journal that covers research conducted at the intersection between ethics and big data.

Planned scientific research grant proposal to SNSF. Finally, based on the rich data resource generated, and the many networking opportunities that emerged during the course of this project, we are convinced to be in an advanced position for a successful grant proposal to SNSF. We will use the generated data resource and propose research in the context of global climate change and urban health. Specifically, we aim to study the impact of heat waves on health outcomes related to heat stress in major European Cities, such as increased mortality rates as documented for example in Paris during the heat wave in 2015. We will investigate how the Twitter sphere could be used as an early surveillance tool in this context to identify emotional stress responses as they emerge in space and time and correlate this information with health outcomes drawn from official health records. We plan for a submission of a full text proposal to SNSF in April 2020.

Glossary

API – Application Programming Interface. In non-technical terms, it can be understood as a 'contract' that says to software developers that if you send a request from a 'client' computer (e.g., a phone, tablet, notebook or desktop) to a 'server' (the computer where the information is stored, in our case the Twitter server) in the specified format you will always get a response in a specified format or initiate a defined action[9].

Bot – Twitter accounts operated by a computer program.

Botometer – A computational tool that detects how likely a Twitter user account is a bot. The tool was developed at the University of Indiana, https://botometer.iuni.iu.edu/#!/faq

DSI – Digital Society initiative at the UZH, https://www.dsi.uzh.ch/de.html

DMP – Data Management Plan that we developed according to the FAIR Data Principles and the SNSF (Swiss National Science Foundation) guidelines.

Encryption – Converting information or data into a code to prevent unauthorized access.

EMOTIVE – An advanced sentiment analysis tool that relies on a semantic model of language to detect fine grained sentiment, specifically a range of eight emotions in a piece of social media text http://emotive.lboro.ac.uk/

FAIR Principles – Data that are Findable, Accessible, Interoperable, and Reusable. The principles were published in the Journal *Scientific Data* under the title 'FAIR Guiding Principles for scientific data management and stewardship' in 2016[10].

Geo-referenced tweet – Tweets that include location metadata.

Geolocation – is an estimation of the real-world geographic location of an object.

Metadata – This is data about data, specifically metadata about a tweet includes information such as the date-time when it was created, how many times the tweet was re-tweeted or favorited, etc.

Rehydration – rehydration process involves sending requests to the official Twitter API /statuses/lookup endpoint[11] with tweet IDs, which are 18-19 long numeric IDs (e.g., 1037658048414842882), uniquely identifying a tweet, where Twitter responds with the tweet and related metadata, when available.

Sentiment Analysis – Usually computational techniques that detect affective language (e.g., positive, negative, neutral) expressed in text, such as a tweet.

[9] Braunstein, Mark L. (2018). Health Informatics on FHIR: How HL7's New API is Transforming Healthcare. Springer. p. 9. ISBN 978-3-319-93414-3.
[10] https://www.go-fair.org/fair-principles/
[11] See documentation – https://developer.twitter.com/en/docs/tweets/post-and-engage/api-reference/get-statuses-lookup.html

Repository – Where data is stored and managed.

R Markdown document – A file format for making code documents with the programming language R which is written in easy-to-write plain text format.

1. Data collection and documentation

1.1 What data will you collect, observe, generate or reuse?

The dataset consists of original geo-referenced Twitter tweets for the time frame between 15.03.2014 to 06.09.2018 for the wider European Region. Each tweet contains geo-location information (geographic coordinates), a time stamp for the time the tweet has been sent, the tweet text, the identified language of the tweet text as identified by Twitter, along with other information made available by the public Twitter API. We also estimated for each English language tweet, whether it contains one of the eight basic emotions: anger, confusion, disgust, fear, happiness, sadness, surprise, and shame and labeled it accordingly. Furthermore, German language tweets found in Switzerland were machine translated into English language and emotion detection was also applied. In addition, for every German tweet geo-located in Switzerland, we calculated the probability of a tweet being produced by a human Twitter user (i.e., a real person) versus a non-human Twitter user (i.e., a bot).

The study region was defined by a bounding box of geographic coordinates (lower left coordinates longitude=-34.402 W, latitude=24.388 N; upper right coordinates longitude=49.935 E, latitude=72.091 N) and includes parts of the Middle East and North Africa, Central Asia, and European Russia.

The tweets stem from 66 countries: Albania, Andorra, Armenia, Austria, Azerbaijan, Belgium, Bulgaria, Bosnia and Herzegovina, Belarus, Switzerland, Cyprus, Czech Republic, Germany, Denmark, Algeria, Egypt, Spain, Estonia, Finland, France, United Kingdom, Georgia, Greece, Croatia, Hungary, Ireland, Iran, Iraq, Iceland, Israel, Italy, Jordan, Kazakhstan, Kosovo, Kuwait, Lebanon, Libya, Liechtenstein, Lithuania, Luxembourg, Latvia, Morocco, Monaco, Moldova, Macedonia, Mali, Malta, Montenegro, Mauritania, Netherlands, Norway, de facto and de jure Palestinian Territories, Poland, Portugal, Romania, Russia, Saudi Arabia, San Marino, Serbia, Slovakia, Slovenia, Sweden, Syria, Tunisia, Turkey, Ukraine and Vatican. However, only parts of contiguous Russia, Kazakhstan, Iran, Saudi Arabia, Algeria, Mali, and Morocco/Western Sahara are included in the study region. Hence, not all tweets from these countries are included in the repository.

1.2 How will the data be collected, observed, and in some cases generated?

The Harvard Centre for Geographic Analysis (CGA) records geo-referenced Twitter data worldwide since 2012. We obtained a subset of CGA's geo-referenced Twitter tweet identifiers (IDs) for the wider European Region for the years 2014-2018 (see specific countries above). These IDs were used to download the full geo-referenced tweets from the official Twitter API (a process called rehydration). However, not all tweet IDs as originally identified by CGA were still publicly available from Twitter user accounts by the time of our download that started in December 2018. If a tweet was deleted or made private by the user before we rehydrated it, it is not included in the repository. Also, a subset of the downloaded tweets was translated into the English language using neural machine translation. This subset included tweets published in Switzerland. Both the translated and original English tweets were subsequently interpreted with EMOTIVE[1], a natural language processing software used

[1] http://emotive.lboro.ac.uk

to detect and measure eight basic emotions within English language text (for specific emotions, see above).

1.3 What documentation and metadata will you provide with the data?
To help secondary users understand and reuse the data, we will create a data dictionary with variable names explained, and units of measurement noted along with brief descriptive statistics (mean, median, range, category, and percentages). Furthermore, we will publish the following metadata on an online website that will also explain the project's aims along with further information on the data sharing policies:

- Details on the data source
- Geographic area and time frame for which the data was collected
- Information about the performed pre-processing steps (e.g. excluding incomplete data entries and bots)
- Analytical steps (EMOTIVE, neural machine translation, geo visualization, Botometer[2]).

2. Ethics, legal and security issues
To ensure compliance with current ethical, legal, and security requirements, our team worked in collaboration with the UZH Department of Data Protection and the data protection officer of the Canton of Zurich between May and June 2019. To ensure conformity of these data and approaches with regard to the general data protection regulation (EU-GDPR) and the Act on Information and Data Protection of the Canton Zurich (IDG), we developed a data sharing and confidentiality agreement document that needs to be signed by all parties involved in working with this data. The document can be found as supplementary material at the end of this DMP. The following paragraphs describe the main ethical, legal, and security issues that can potentially emerge from handling social media data, such as from Twitter, and the strategies we pursue for limiting any possible risks for the Twitter users involved.

2.1 How will ethical issues be addressed and handled?
Privacy
Current ethical guidelines suggest that collecting data from a public space is appropriate when people could "reasonably expect to be observed by others". Twitter is such a public space for online communication and information sharing. Most content that is shared on Twitter is publicly accessible via the Twitter API and/or via data resellers. Furthermore, permissions to access certain information of a Twitter account are not granted by default; users can choose on a per-application basis whether to provide this access, and can control all the applications authorized on their account. Providing access may be interpreted as giving up a certain degree of privacy and making information available to third parties. Finally, Twitter clearly states in its Privacy Policy, that Twitter is public and that anyone in the world can immediately access what Twitter users decide to publicly share on the platform and that Twitter users should not upload or tweet anything that they are not comfortable with publicly sharing.

Nevertheless, one of our prime goals is to safeguard the privacy of the tweet authors. To do this, we will:

[2] https://botometer.iuni.iu.edu/#!/publications

- carefully assess whether the Twitter users involved actually intended to make their information public in the light of their privacy settings or limited audience to which the data were made available. Therefore, we only rehydrated those tweets that are available via the Twitter API. This procedure effectively provided for a recent assessment of whether consent had been withdrawn by the user (by deleting the tweet from back in 2014 to 2018 or switching a profile to private in the meantime).
- guarantee encrypted storage of all retrieved data using the AES 256-bit encryption standard;
- avoid the publication of re-tweets, screen names or any mentions within tweets or other details that can be used to identify a subject
- before publication, aggregate tweets to larger geographic areas, that is, areas which are sufficiently large and populous to make inference of the specific tweet creator's identity unfeasible
- share only agreed subsets of the original, full dataset with third parties (i.e. researchers external to the M-HEALTH project)
- share data with third parties only if:
    - The data confidentiality agreement has been signed by all parties,
    - The data management plan has been submitted and reviewed by the UZH Department of Data Protection, and
    - A local Institutional Review Board has provided clearance of the planned project
- ensure oversight on the application of the data confidentiality agreement.

Full anonymization will not be possible due to the nature of Twitter data. For this reason, we will retain data on an individual tweet level and only publish data on an aggregated level, that is, tweets at some geographic area level that does not allow for individual user identification.

Furthermore, we will keep local copies of tweets that contain detected emotions separate from the original Twitter tweets.

Informed consent

The value of tweets as a data source relies on its size. We will be working with a large dataset that does not make it feasible to obtain informed consent from all users that are part of it. Furthermore, it may not be possible to reach all Twitter users involved, because when they are approached for informed consent via a tweet or a direct message, they may no longer be maintaining their account, simply not reply or take a long time to reply which would fall beyond the realistic temporal scope of this project. Due to the unfeasibility of obtaining informed consent, for this project, we will describe tweets at an aggregated level, that is, tweets at some geographic area level that does not allow for individual user identification and provide this meta information on a website.

Legal concerns

From a legal point of view, the Twitter developer agreement (Effective: May 25, 2018.) allows extraction of the data for research purposes, unless it "*isolates a group of individuals or any single individual for any unlawful or discriminatory purpose or in a manner that would be*

inconsistent with our users' reasonable expectations of privacy" (https://developer.twitter.com/en/developer-terms/agreement-and-policy.html). To ensure compliance with Twitter developer agreement, any actual and subsequent use of the data must be consistent with Twitter users' reasonable expectations of privacy and must prevent any unlawful or discriminatory usage of the data as outlined in the data sharing agreement that any researchers interested in the data will have to sign.
Twitter's General guidelines and policy state "*We make public Tweets and replies available to developers, and allow developers to post Tweets via our API. Developers can access Tweets by searching for specific keywords, or requesting a sample of Tweets from specific accounts*" (https://help.twitter.com/en/rules-and-policies/twitter-api). This makes the process of data extraction fully lawful. Regarding data sharing, Twitter further states "*If you need to share Twitter content you obtained via the Twitter APIs with another party, the best way to do so is by sharing Tweet IDs, Direct Message IDs, and/or User IDs, which the end user of the content can then rehydrate (i.e. request the full Tweet, user, or Direct Message content) using the Twitter APIs. This helps ensure that end users of Twitter content always get the most current information directly from us*" (https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases).
We fully comply with these requirements and only rehydrated those tweets from the Twitter API.

2.2 How will data access and security be managed?
A log file will help us track all accesses to the data repository. We will separate identifying data (Twitter IDs, Twitter screen names, etc.) from subsets for any analyses. All digital documents will be stored on a password protected and encrypted storage drive. Paper data will be securely locked away and securely destroyed once they are no longer needed.
The individual persons who have authority to full access of the data subjects (Twitter users) are: Dr. Oliver Gruebner, Tamar Edry, and Ivor Mardesic of the Epidemiology, Biostatistics, and Prevention Institute and the Department of Geography of the University of Zurich; Prof. Emiliano Albanese and Dr. Marta Fadda of the Faculty of Biomedical Sciences of the Università della Svizzera italiana (USI); Dr. Martin Sykora and Dr. Suzanne Elayan of the Centre for Information Management (CIM), School of Business and Economics of Loughborough University. Researchers interested in accessing to the data will have to obtain permission and sign a data sharing and confidentiality agreement with the principal investigator of the project (see supplementary material at the end of the document).

2.3 How will you handle copyright and Intellectual Property Rights issues?
To prevent any unauthorized use of a copyrighted image such as a profile or header photo, copyrighted video or image uploaded through Twitter media hosting services, or tweets containing links to unknown URLs, we will not publish this type of data, and we will make this requirement explicit in the data confidentiality agreement.

3. Data storage and preservation
3.1 How will your data be stored and backed-up during the research?
All health-related personal data are password protected, stored on a local and encrypted data drive (AES 256-bit encryption standard), and are regularly backed-up. The back-up files are additionally encrypted and copied to an UZH internal network drive that is regularly backed-

up by the UZH Department of Geography IT service. The double back-up procedure ensures that no data will be lost if the local drive will break.

We will process all original data with R. All steps of data pre-processing are therefore documented and any subsequent changes are tracked. Access is only possible with a password. In some cases, aggregated data will also be processed with ESRI's ArcGIS to produce specialized maps.

3.2 What is your data preservation plan?

All data will be retained after the completion of this project for the purpose of achieving the project goal, that is, to create a repository. We aim at subsequent projects that apply these data in an analytical way. As detailed above, all subsequent projects and analyses will seek new approvals from data security and ethics committees.

After completion of any subsequent analyses with these data, the data will be retained and archived for the purpose of good scientific practice and reproducibility. All security, data protection and confidentiality measures that are detailed in this document will be applied to the archived data. The data will be stored as comma separated value (CSV) files along with a data dictionary applying respective security and data confidentiality standards as detailed above.

4. Data sharing and reuse

4.1 How and where will the data be shared?

An online web mapping application will serve as an information tool to explore which data is available from the repository through documentation and meta-information (the mapping application can be found here: https://givauzh.shinyapps.io/tweets_app/). The raw data itself will be kept on secured local drives at the University of Zürich, as detailed above (section 3.1). The data will be shared with third party researchers only for research purposes and only after detailed review of the planned research. Third party researchers will therefore have to submit 1) a brief research proposal of about two pages, in which the purpose and methodology are detailed, 2) a DMP, in which the data security and confidentiality measures are detailed (in accordance with this DMP), 3) the signed data confidentiality agreement (attached to this DMP), and 4) clearance of a local Institutional Review Board. The application will be reviewed by the principal investigator of this project and approval has to be sought by the UZH Department of Data Protection. After final agreement by all involved parties, only the desired subset of the data for which the applicants opted-in may be shared with third party researchers. Furthermore, it will be the duty of the third party to check via the official Twitter API whether tweets have been deleted by the user (the original content provider) in the meantime, delete these tweets accordingly on the local drives, and report those back to the M-HEALTH project for deletion.

4.2 Are there any necessary limitations to protect sensitive data?

To protect Twitter users' privacy, we will avoid the publication of re-tweets, screen names, or any mentions within tweets other details that can be used to identify a subject, by reducing the tweet mention/retweet data to a binary value; we will aggregate geo-location information to larger geographic areas; and we will share only portions of the original, full dataset with third parties only after a review process (see above) and a data sharing and confidentiality agreement has been signed.

4.3 Choosing digital repositories that are conform to the FAIR Data Principles
The data repository will conform to the FAIR Data Principles (Findable, Accessible (after above conditions are met), Interoperable and Reusable, https://www.go-fair.org/fair-principles/).

4.4 Choosing digital repositories maintained by a non-profit organization
This option will not be applicable since we have to comply with data confidentiality measures. Therefore, the data remains at a local drive and only metainformation is accessible online. The digital repository is maintained at the University of Zurich, a public higher education institution of the Canton of Zurich.

Supplementary material

Data sharing and confidentiality agreement

This agreement governs the confidentiality of data used within the "*Social Media Mental Health Surveillance M-HEALTH"* project. Project stakeholders have a legal interest in ensuring that confidential information pertaining to the social media data used during this project is not disclosed to unauthorized third parties and that privacy and security practices are followed. This confidentiality agreement constitutes the broad framework to which all project stakeholders will abide. Further documents amending specific cases may be set forth in the future by the principal researchers of the project (*Oliver Gruebner*, *Emiliano Albanese*, and *Milo Puhan*).

These confidentiality and privacy standards are mandated by our commitment to the highest ethical standards in science and by our responsibilities to abide with Swiss and EU regulations, Twitter policies, and the rights to privacy of any identifiable individual record in the data (person, household, organization, business etc.). All data are provided *"as is."* The University of Zurich provides no certifications or warranties, explicit or implicit, regarding the data's accuracy, completeness, merchantability or use. Privacy is personality protection. Thus, not only the data as such, but also the person to whom this data concerns is to be protected. In order to avoid legal and ethical issues when interacting with the social media data from the "*M-HEALTH"* project, this agreement will set forth several binding protocols.

I understand that my use of **Social Media Data** from the study "M-HEALTH" requires that I follow the protocols listed below.

I will:

1. Use the dataset(s) only for the purposes of this project, i.e. research;
2. Store all data files on encrypted, password protected devices (at least AES 256-bit encryption standard);
3. Store any back-ups of the data files on encrypted, password protected devices (at least AES 256-bit encryption standard);
4. Conduct only the analyses that have been approved by the principal researchers;
5. Preserve the confidentiality of information pertaining to identifiable individual records (person, household, organization, business etc.) that are recorded in the dataset(s);

6. Ensure with utmost diligence that no identifiable information is included in published or unpublished reports, articles, or other written material that appears in print or in electronic form;
7. Provide the "*M-HEALTH"* project with copies of publications and other research reports based on these dataset(s);
8. Ensure personal data is protected regardless of whether it is processed manually or electronically;
9. Abide by any further conditions and documents as set forth by the principal researchers of the project;
10. Delete all files after five years, latest.

I will not:

1. Make copies of the data other than security and back-up copies for my own use;
2. Allow parties outside the "*M-HEALTH"* project access to the dataset(s);
3. Attempt to link the data to other dataset(s) for which I have not obtained prior approval by the UZH data protection office and a local ethical review board;
4. Attempt to identify any individual record (person, household, organization, business etc.) in the dataset(s);
5. Release or publish any information or results that identify any individual record or may lead to the identification of any individual record;
6. Use any other data in this project for which I have not signed a confidentiality agreement approved by the principal investigators of this project;
7. Publish copyrighted images such as a profile or header photo, copyrighted video or image uploaded through Twitter media hosting services, or tweets containing links to unknown URLs.

I acknowledge that I have read all of the above conditions, that I understand them, and that I am personally accountable for correct and responsible use of the data and data access system. Furthermore, I understand that my failure to comply with these conditions will result in the withdrawal of my access to the dataset and any other sanctions that may be determined by the University of Zurich and/or may be specified by law.

Name: _____

Signature: _____

Date: _____

Please provide contact information below, so that we can contact you if new data become available or if revisions are made to the data:

E-mail address: _____

Telephone: _____