International Journal of Data Science and Analytics
https://doi.org/10.1007/s41060-020-00224-2

REVIEW

Measuring objective and subjective well-being: dimensions and data sources

Vasiliki Voukelatou[1] · Lorenzo Gabrielli[2] · Ioanna Miliou[3] · Stefano Cresci[4] · Rajesh Sharma[5] · Maurizio Tesconi[4] · Luca Pappalardo[2]

Received: 18 July 2019 / Accepted: 8 May 2020

Abstract

Well-being is an important value for people's lives, and it could be considered as an index of societal progress. Researchers have suggested two main approaches for the overall measurement of well-being, the objective and the subjective well-being. Both approaches, as well as their relevant dimensions, have been traditionally captured with surveys. During the last decades, new data sources have been suggested as an alternative or complement to traditional data. This paper aims to present the theoretical background of well-being, by distinguishing between objective and subjective approaches, their relevant dimensions, the new data sources used for their measurement and relevant studies. We also intend to shed light on still barely unexplored dimensions and data sources that could potentially contribute as a key for public policing and social development.

Keywords Well-being · Objective well-being · Subjective well-being · Well-being dimensions · New data sources · Data science for social good · Artificial intelligence for social good

1 Introduction

Economists and policy-makers have traditionally considered gross domestic product (GDP) as a good indicator of well-being in society, mainly because it is strongly linked with the standard of living indicators [1]. However, GDP has been criticized as a weak indicator of well-being and, therefore, a misleading tool for public policies [2]. The Stiglitz Commission [3] in 2009 observed that other statistical tools should be used, complementary to GDP, for the measurement of well-being. Therefore, considering that well-being is difficult to be captured only with GDP, researchers with various backgrounds, from economists to psychologists, suggested two main approaches to measuring the overall well-being; objective well-being and subjective well-being.

Defining objective well-being has always been considered a challenging task, and therefore researchers have focused on exploring its dimensions rather than its definition [4,5]. It is due to its objective nature that one could claim that objective well-being could be measured in terms of GDP. However, it must reflect both people's material living conditions and the quality of their lives. In fact, the Organisation for Economic Co-operation and Development (OECD) [6], the United Nations Development Programme (UNDP) [7] and the Italian Statistics Bureau (ISTAT) [8] have identified six major objective and observable dimensions for its measurement: *health, job opportunities, socioeconomic development, environment, safety, and politics*. All these dimensions together represent the objective well-being, which is assessed through the extent

✉ Luca Pappalardo
luca.pappalardo@isti.cnr.it; info@sobigdata.eu

Vasiliki Voukelatou
vasiliki.voukelatou@sns.it

Lorenzo Gabrielli
lorenzo.gabrielli@isti.cnr.it

Ioanna Miliou
ioanna.miliou@for.unipi.it

Stefano Cresci
stefano.cresci@iit.cnr.it

Rajesh Sharma
rajesh.sharma@ut.ee

Maurizio Tesconi
maurizio.tesconi@iit.cnr.it

[1] Scuola Normale Superiore and ISTI-CNR, Pisa, Italy

[2] ISTI-CNR, Pisa, Italy

[3] University of Pisa, Pisa, Italy

[4] IIT-CNR, Pisa, Italy

[5] University of Tartu, Tartu, Estonia

Published online: 29 June 2020

to which these "needs" are satisfied. The objective approach investigates the objective dimensions of a good life, whereas the subjective approach examines people's subjective evaluations of their own lives. In 2013, the OECD [9] recognized the importance of taking into consideration people's perceived well-being, labeled as subjective well-being when investigating the overall well-being. Subjective well-being, also called happiness, has been defined by Veenhoven [10], as the degree to which an individual assesses the overall quality of her life-as-a-whole favorably. This might as well be different as compared to GDP, which cannot be representative of societal happiness. Indeed, GDP explains only a small proportion of its variations on humans [11], and it might be different from people's perceptions of their well-being [12]. Therefore, subjective well-being has been traditionally captured through studies based on data collected by self-reports. These studies highlight five main dimensions of subjective well-being: the *role of human genes*, which seem to be fairly heritable [13–21], *universal needs*, meaning basic and psychological needs [22–24], *social environment*, such as education and health [25–29], *economic environment*, including a lot of research on income [30–34], and *political environment*, such as democracy and political freedom [35,36].

Traditionally, both objective and subjective well-being are measured through surveys of household income and consumption [37]. Although these surveys have been considered accurate and valid, they bring some considerable disadvantages. For example, they cannot provide constant updates of well-being to policy-makers, and they have high costs to be conducted, making it difficult for many developing countries to estimate well-being frequently. The last few years have witnessed a drastic change in the approaches used to measure well-being. Researchers of different disciplines propose several innovative data sources and methods, which could potentially overcome the limitations of the traditional methods for the individual and collective well-being measurement, both objective and subjective.

To support research in this direction, the European project SoBigData [38] has created a virtual environment within a research infrastructure that provides theoretical knowledge, data, and innovative methods to scholars that want to address challenging questions involving both objective and subjective well-being.

Therefore, in line with the purposes mentioned above and the support of SoBigData, the aim of this paper is to provide the theoretical background on objective and subjective well-being, including their relevant dimensions. Additionally, the article seeks to present to researchers the new data sources used for capturing well-being, as well as discuss indicative existing studies.

We believe that this study offers great value to the scientific community and especially to researchers interested in "Data Science for Social Good" (DS4SG) or similarly "Artificial Intelligence for Social Good" (AI4SG) [39], since it could work as a reference point for adequate measurement of well-being with the use of innovative data sources and tools. In particular, at this critical moment that the global society is under financial and political crisis and instability, policy-makers need frequent updates of well-being. This could facilitate them to react on time on applying the right policies to prevent detrimental societal effects and contribute effectively to societal progress.

The remainder of this paper is organized as follows: It is divided into two main sections, as suggested from the literature, i.e. objective and subjective well-being. In particular, Sect. 2 is dedicated to objective well-being and Sect. 3 is dedicated to subjective well-being. For both sections, we provide a theoretical background on objective and subjective well-being and their dimensions respectively. We then provide the data sources used for monitoring well-being. Besides, we present essential studies on well-being; to present them in an organized flow, we categorize the presentation of the studies by matching each well-being dimension separately with each data source. Finally, in Sect. 4, we provide a discussion on the study, highlighting the opportunities for future research on well-being.

2 Measuring objective well-being

Suggesting a single definition of objective well-being is a substantial challenge, mainly due to its multi-dimensionality. Therefore, researchers have focused on carefully specifying its objectively measurable dimensions [4,5]. Objective well-being is traditionally captured through surveys, such as household income and consumption surveys [37]. However, usually, such surveys are very costly and time-consuming [40], making it difficult for many countries and global institutes to update their estimates frequently. Therefore, the last few years have witnessed a change in the way of measuring objective well-being. In particular, researchers of various disciplines propose several methodologies to measure individual and collective objective well-being, based on a combination of new data sources and traditional surveys [41–44]. The United Nations also stimulate this change of studying well-being in two recent reports, where the usage of new, mostly big, data sources, is encouraged for the investigation of patterns of phenomena related to people's health and well-being [45,46].

2.1 The dimensions of objective well-being

During the last years, public institutions and non-governmental companies have worked on identifying dimensions that are considered essential for the improvement of the societal well-being and its comparison between countries and years.

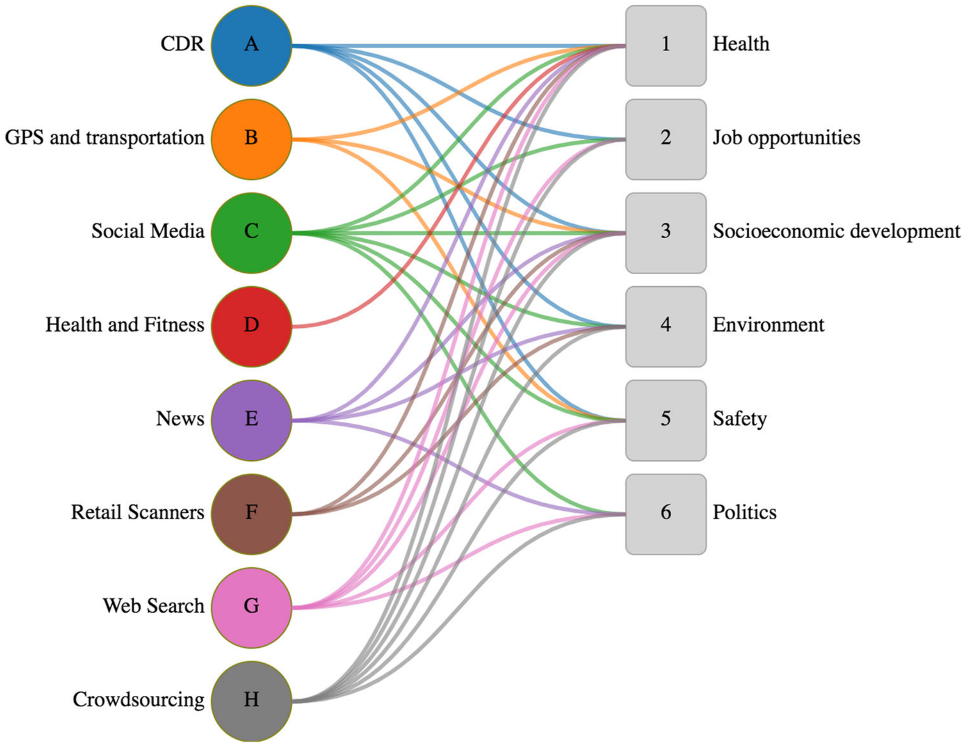

Fig. 1 The figure relates the sources of data (left) with the dimensions of the objective well-being (right)

For example, the Organisation for Economic Co-operation and Development (OECD) has identified 11 essential topics labeled as OECD well-being framework [6]; the United Nations Development Programme (UNDP) has identified 17 sustainable development goals, labeled as SDGs [7]; and the Italian Statistics Bureau (ISTAT) has created an ambitious project named "Benessere Equo e Sostenibile" (BES) that stands for "Fair and Sustainable Well-being" [8]. From the initiatives mentioned above, it is evident that for different institutions, well-being dimensions might be different, sometimes vague, and statistically hard to be captured. Therefore, based on the aforementioned official authorities, we suggest the following concrete and measurable dimensions of well-being (Fig. 1).

2.1.1 Health

Health status represents an essential factor for people's well-being, as shown by the WHO Commission on Macroeconomics and Health in 2001 at global level [47], and by the Lisbon Strategy for Growth and Jobs in 2000 [48]. Health brings together many other benefits, from job opportunities to social relationships, from reduced health care costs to an increased life expectancy. Indeed, there have been remarkable gains in life expectancy over the past 50 years in OECD countries [49], due to the health care spending growth, lifestyle, educational, and environmental changes. Chronic (non-communicable) diseases, such as cancer, diabetes, and chronic respiratory conditions, are nowadays the primary disability and mortality factors in OECD countries. Fortunately, some indicators can help prevent the diseases mentioned above. For example, the number of people who are driving carefully, who are non-smokers or who do not drink a large amount of alcohol, are risk-indicators, which, if taken into consideration, could contribute to an improvement in the health status of a territory.

2.1.2 Job opportunities

This is a crucial dimension of well-being since it has obvious economic and societal benefits, contributing to people's health and societal, political, and economic stability. The job opportunities dimension is composed of three main determinants: employment rate, quality of work, and work–life balance. The employment rate is a crucial aspect since individuals in countries with a high level of employment, are well connected in society. In particular, it is a proxy used by policy-makers to avoid poverty and social exclusion. The second determinant is the quality of work, in terms of objective working stability, retribution, skills, and safety at work, which might show some differences between different working environments. Moreover, work-life balance is the determinant that mainly aims to capture the balance between work and life. In the OECD countries, a full-time worker devotes 62% of the day on average (15 hours) on personal care (e.g., eating, sleeping) and leisure (e.g., socializing with friends and family, hobbies) [50]. This determinant is mainly created to capture women's work-life balance. Indeed, the quality of a country's employment is measured by the balance women have between family care and paid work.

2.1.3 Socioeconomic development

While socioeconomic indicators alone do not suffice to represent societal well-being, it cannot be doubted that they positively influence it. The variables that contribute to its measurement are income, wealth, consumption expenditure, housing conditions, and possession of consumer durables, and it can implicitly influence access to university, health care, and more. In particular, the Organization for Economic Co-operation and Development (OECD) [6] and the Italian Statistics Bureau (ISTAT) [8] suggest two main determinants that constitute the overall economic well-being: available income and wealth, and consumption expenditure.

In a market economy, income measures the purchasing capacity of individuals, and it is thus an essential predictor of economic well-being. Wealth, on the other hand, takes into account savings, monetary gold, stocks, securities, and loans [51]. Therefore, wealth could be considered an essential source of revenue, which could make people less vulnerable to difficult economic situations that might affect their life.

Additionally, consumption expenditure is a direct estimate of the goods and services that contribute to determining the living conditions of individuals. Unlike income, consumption expenditure can contribute to making interpersonal comparisons, since it captures whether each individual can acquire her desired goods and services.

2.1.4 Environment

A healthy natural environment is essential for all individuals' well-being in society. Clean water, clear air, and uncontaminated food are examples of goods that can only be possible in an environmental context where humans' productive and social activities are made with respect to the environment and its natural resources. For the reasons mentioned above and due to the recent environmental crisis, the United Nations set sustainable environmental goals [7], such as Clean Water and Sanitation, Climate Action, and more. Similarly, ISTAT [8] suggests five determinants for describing the interactions between society and the environment that are connected. These determinants are quality of the water, quality of the air, quality of the soil and the land, biodiversity, and matter, energy, and climate change. Finally, the "OECD Environmental Outlook to 2050" projects the number of premature deaths associated with exposure to PM10 and PM2.5 to increase from just over 1 million worldwide in 2000 to about 3.5 million in 2050 [52]. Therefore, the more these determinants are taken into consideration by policy-makers and by citizens' activities, the more the citizens can contribute to radical changes for the protection of societal well-being.

2.1.5 Safety

It includes the risk of people being physically assaulted, falling victims, and suffering from other crimes, such as economic loss, physical damage, and psychological posttraumas stress. Reducing violent crime, sex trafficking, forced labor, and child abuse are clear global goals, as suggested by the United Nations [7]. Besides, the Italian BES project [8] suggests that safety is characterized by two determinants: criminality and violence.

Criminality is one of the most common security threats in developed and emerging countries, and it has both a direct and indirect impact on people. It directly influences individuals' health (physical and mental) and economic situation. According to the latest OECD data, the average homicide rate in the OECD is 3.6 murders per 100,000 inhabitants [53]. Indirectly, criminality has an impact on non-victims' well-being when being on victims' social network or by news spread on (social) media.

Another determinant is violence suffered inside and outside the family and it has both a direct and indirect impact on people. In particular, victims suffer from the direct effects, which can last for long periods, if not for the whole life, depending on individuals' ability to manage their daily life, medical expenses, dependence on others, and capacity to achieve happiness. Indirectly, it causes insecurity and anxiety, which brings difficulties in their daily activities [54].

2.1.6 Politics

This dimension is also essential for objective well-being. Today, due to the economic crisis, more than ever, citizens demand greater transparency from their Governments and the Public Institutions. Fair civic and political participation, as well as transparency, do not only contribute directly to well-being but also indirectly since they allow greater efficiency of public policies, a lower cost of transactions, and the minimization of the risk of fraud. Therefore, two determinants fall under this category, which are associated with the Public Sphere as a driver of the individuals' well-being, on either local or national level: civic and political engagement, and trust and social cohesion. Voter turnout is the best existing means of measuring civic and political engagement, and is measured as the percentage of the registered population that voted during elections. According to OECD data, voter turnout, is averaged 69% in OECD countries, which shows that not everyone exercises the voting right [55]. Regarding trust and social cohesion, OECD suggests public engagement (e.g., stakeholder engagement) for developing regulations [55]. If citizens have the possibility to participate in the development of laws and regulations, it is more likely that they will trust the government institutions and they will comply with the societal rules.

Table 1 Pros and cons for each data source used for the measurement of objective well-being

Data source	Pros	Cons
CDRs	Temporal and social dimensions, world wide diffusion, repeatability	Not publicly available, sparsity, geographically imprecise
GPS and transportation	Coverage of rural areas, unbiased and classified, real-time monitoring	Privacy issues, indoor spatial inaccuracy
Social Media	Measuring social dynamics, publicly available	Privacy issues, overrepresentation, social desirability bias
Health and Fitness	Cost effective, applicable for multiple studies, prediction of near-term risk of events	Not publicly available, not necessarily representative of the population, limited time slots
News	Variety of subject domains, range of targets, 24/h updated, archived historical news	Gatekeeping bias, coverage bias, statement bias
Retail Scanners	Modeling of dynamic household behavior, control time-invariant characteristics, long term coverage, quality improvement of HICP	Dependency on retailer's permission, legal constraints
Web Search	Publicly available, speed, convenience, flexibility, ease of analysis	Population size varies across domains, hard identifying relevant queries
Crowdsourcing	Large number of data, speed, relative low cost	Risk of low-quality results, trade-off between quality and cost

2.2 Data sources for monitoring the dimensions of objective well-being

Figure 1 describes the new data sources (left) that have been used to estimate one or more dimensions of objective well-being (right). The presence of a link in Fig. 1 between a data source and a dimension indicates that there are papers in the literature on monitoring that dimension with that data source. In this section, we describe, for each data source, its features (e.g., the process of data collection, its biases and limitations) and the main works in the literature that use it to measure several dimensions of objective well-being. Table 1 provides a summary of the data sources used, highlighting the pros and cons of each one. We refer to a link between a data source and a dimension using a letter-number notation. For example, B3 indicates the link between GPS data (B) and socioeconomic development (3).

Table 2 Example of Call Detail Records (CDRs). Every time a user makes a call, a record is created with timestamp, the phone tower serving the call, the caller identifier and the callee identifier (a). For each tower, the latitude and longitude coordinates are available to map the tower on the territory (b)

(a) Timestamp	Tower	Caller	Callee
2007/09/10 23:34	36	4F80460	4F80331
2007/10/10 01:12	36	2B01359	9H80125
2007/10/10 01:43	38	2B19935	6W1199
⋮	⋮	⋮	⋮

(b) Tower	Latitude	Longitude
36	49.54	3.64
37	48.28	1.258
38	48.22	-1.52
⋮	⋮	⋮

2.2.1 CDRs

Many works in the literature are based on the analysis of mobile phone data, the so-called CDRs (Call Detail Records) of calling and texting activity of users, because they guarantee the repeatability of experiments in different countries and on different scales given the worldwide diffusion of mobile phones [56].

CDRs collect geographical, temporal, and interaction information on mobile phone use [57–62], hence providing a comprehensive picture of human behavior at a societal scale. Each time an individual makes a call, the mobile phone operator registers the connection between the caller and the callee, the duration of the call, and the coordinates of the phone tower communicating with the served phone. Table 2 illustrates an example of the structure of CDRs.

Note that CDRs suffer from different types of bias [63,64]. For example, the position of a user is known at the granularity level of phone towers, and only when they make a phone call. Moreover, phone calls are sparse in time, i.e., the time between consecutive calls follows a heavy tail distribution [65,66]. In other words, since users are inactive most of their time, CDRs allow reconstructing only a subset of a user's behavior.

CDRs are used to monitor several dimensions of well-being, notably health (A1), job opportunities (A2), socioeconomic development (A3), environment (A4), and safety (A5).

CDRs provide one of today's most exciting opportunities to study human mobility and its influence on disease dynamics (A1). Many researchers use mobile phone data for public health, as the analysis of individual and population mobility patterns is more objective and with finer spatiotemporal resolution in comparison to traditional methods. Furthermore, mobile network data can also provide insights into human behavior that can support the assessment and monitoring of the health of specific communities at risk, thus paving the way toward improved health promotion and prevention [67]. Taking into consideration that the spatiotemporal evolution of human mobility and the related fluctuations of population density are essential drivers of disease outbreaks, Finger et al. [68] use CDRs to track the cholera outbreak in 2005 in Senegal. Findings show that a mass gathering taking place during the initial phase of the outbreak has an essential impact on the course of the disease. Besides, Kafsi et al. [69] contribute to the fight against epidemics of infectious diseases using CDRs provided by France Telecom-Orange. They use 2.5 billion calls made by 5 million users in the Ivory Coast, recorded over 5 months, from December 2011 to April 2012, to study and model behavioral patterns of the affected population and propose several strategies for personalized behavioral recommendations to reduce the infections. Lima et al. [70] use the same data set to build a model that describes how diseases circulate the country as people move between regions, and they enhance the model with a concurrent process of relevant information spreading. This process corresponds to people disseminating disease prevention information, e.g., hygiene practices, vaccination campaign notices, and others, within their social network. Finally, Madan et al. [71] use CDRs and mobile phone-based co-location sensing to measure characteristic behavior changes in symptomatic individuals. These behavior changes are reflected in their total communication, interactions with respect to time of day, diversity, and entropy of face-to-face interactions and movement. Using these extracted mobile features, they manage to predict the health status of an individual, without having actual health measurements from the subject.

Besides, researchers use CDRs to study job opportunities (A2). Pappalardo et al. [72] use CDRs to study the link between human mobility and the employment rate of French cities, finding a strong correlation between measures of mobility entropy and the unemployment rate in urban environments. Toole et al. [73] show that changes in the calling behaviors of individuals, aggregated at regional level, can improve forecasts of macro unemployment rates. Sunds et al. [74], use CDRs to create a model which predicts unemployment with a 70.4% of accuracy. They also provide promising support to the collection of data for populations in developing countries, which are often under-represented in official surveys.

Most of researchers use CDRs to investigate socioeconomic development (A3). A seminal work analyzes landline calls and a nationwide mobile phone data set to show that, in the UK, regional communication diversity is positively associated with a socioeconomic ranking [75]. Other works address the issue of mapping poverty [76] and other socioeconomic determinants [77] with mobile phone communication data, combined with airtime credit purchases data in the Ivory Coast [78]. Blumenstock et al. [79,80] show preliminary evidence of a relationship between individual wealth and the history of mobile phone transactions. Frias-Martinez et al. [81–84] analyze the relationship between human mobility and the socioeconomic status of urban zones, presenting which mobility indicators correlate best with socioeconomic levels and building a model to predict the socioeconomic level from mobile phone traces. Pappalardo et al. [85] analyze mobile phone data and extract meaningful mobility measures for cities, discovering an interesting correlation between human mobility aspects and socioeconomic determinants. Lotero et al. [86] analyze the architecture of urban mobility networks in two Latin-American cities from the multiplex perspective. They discover that the socioeconomic characteristics of the population have an extraordinary impact on the layer organization of these multiplex systems. In a successive work, Lotero et al. [86] analyze urban mobility in Colombia representing cities by mobility networks. They encode the origin-destination trips performed by a subset of the population corresponding to a particular socioeconomic status and they show that spatial and temporal patterns vary across these socioeconomic groups. Amini et al. [87] use mobile phone data to compare the human mobility patterns of a developing country (the Ivory Coast) and a developed country (Portugal). They show that cultural diversity in developing regions can present challenges to mobility models defined in less culturally diverse regions. Smith-Clarke et al. [88] analyze the aggregated mobile phone data of two developing countries and extract features that are strongly correlated with poverty indexes derived from official statistics census data.

Moreover, researchers use CDRs to monitor the quality of the environment and its impact on people's lives (A4). For example, Picornell et al. [89] evaluate the population exposure to NO_2 on a research published recently. They use CDRs from one of the three most important Spanish mobile phone network operators (MNOs), with around 30% market share. The analysis is conducted for the capital of Spain, Madrid, for the 17th of November 2014, as a typical day in terms of population mobility and NO_2 levels. Comparing the results with traditional census-based methods, they demonstrate relevant discrepancies at disaggregated levels and underline the importance of integrating CDRs data for the evaluation of population exposure to NO_2. Lu et al. [90] study people's behavior affected by climate stress. In particular, by exploring the individuals' behavioral response to the Cyclone Mahasen, which struck Bangladesh in May 2013, they find out that anomalous patterns of mobility and calling frequency correlate with rainfall intensity, showing the affected regions and when the storm moves. Lu and Bengtsson [91,92] analyze the movement of 1.9 million mobile phone users before and after the 2010 Haiti earthquake, and they show that CDRs can be a valid data source for estimates of population movements during disasters. Wilson et al. [93] build a tool within nine days of the Nepal earthquake of 2015, to provide spatiotemporally detailed estimates of population displacements from CDRs based on movements of 12 million mobile phones users. Nyarku et al. [94] use CDRs to explore whether mobile phones could be reliably used to monitor individual exposure to selected air pollutants when moving between indoor and outdoor microenvironments. In particular, data are collected from two BROAD life mobile phones, which are equipped with sensors for direct measurements of air pollutants. The two phones bring similar results, both for particles and formaldehyde, making them potentially suitable for applications in polluted environments, even if there seem to be some exceptions where the readings of the two phones do not correspond well to each other. Liu et al. [95] map personal trajectories using mobiles in an urban environment to assess the impact of traffic-related air pollution in society. They estimate traffic pollution exposure to individuals based on the exposure along the individual human trajectories in the estimated pollution concentration fields by utilizing modeling tools and manage to identify trajectory patterns of particularly exposed human groups. In addition, Decuyper et al. [96] use CDRs to study food security indicators finding a strong correlation between the consumption of vegetables rich in vitamins and airtime purchase.

Other studies focus on the safety dimension (A5). Bogomolov et al. [97] use CDRs for 3 weeks from the 9th to the 15th of December 2012 , and from the 23rd December 2012 to the 5th of January 2013, in combination with demographic data from December 2012 to January 2013, to predict crime in the city of London. Experimental results show 70% of accuracy in predicting whether an area could be a crime hotspot or not. Similarly, Ferrara et al. [98] study criminal networks to detect and characterize criminal organizations in networks reconstructed from the CDRs. They also introduce an expert system to support law enforcement agencies in unveiling the underlying structure of criminal networks.

2.2.2 GPS and transportation data

Since the 1990s, Global Positioning Systems (GPS) have been used for tracking the movements of the individuals [99–102]. In particular, GPS data provide time and location coordinates information, which can be used to link locations with environments and to calculate the speed of movements [103]. For insurance reasons, some vehicles have a black box installed. The device records the position of the vehicle at regular intervals and sends it to the database. Table 3 illustrates an example of the structure of GPS records.

Table 3 Example of GPS records

Vid	Timestamp	Latitude	Longitude
63	2014-06-18 06:31:24	43.557703	10.337913
63	2014-06-18 06:31:26	43.557725	10.33794
63	2014-06-18 06:31:27	43.557735	10.337955
⋮	⋮	⋮	⋮

The collected GPS data consist of the sequence of space-time detections of vehicles on which the positioning device is installed. Every time a vehicle switches on, a record is created consisting of the vehicle identifier, timestamp, the latitude and longitude coordinates

GPS data can also cover rural areas, as opposed to other data, mostly collected among citizens of urban areas [104]. Comparing to the traditional ways of measuring mobility, usually by self-reports assessed with questionnaires, GPS does not bring any biases and misclassification, [104,105], as it eliminates the social desirability usually brought by self-report participants [106,107]. Another advantage of GPS data is that they provide real-time monitoring. However, while there are studies based on GPS data covering hundreds of thousands of individuals [108] most of the GPS studies are conducted with fewer than 300 participants [104,109], usually due to privacy issues. Apart from this drawback, when a GPS is used indoors, the spatial accuracy of the measurements is fairly detected [110], which creates problems in specific fields, such as on epidemiology research.

GPS data are used to explore several dimensions of objective well-being, notably health (B1), socioeconomic development (B3), and safety (B5).

Health (B1) exploration has also attracted the interest of researchers. For example, Saelens et al. [111] track the movements of an individual through GPS devices and bring to the surface growing evidence that transit users are more physically active than non-transit users, which could potentially lead to the health improvement of the first ones. Similarly, Rundle et al. [112] explore health in terms of physical activity, and conclude that neighborhood walkability influences other residents' choice of space utility and is also associated with higher weekly physical activity. Additionally, Sadler et al. [113] use GPS data to understand children's exposure to junk food in Canada and compare the results to a validated food environment database. They demonstrate that official results underestimate exposure to junk food up to 68%, which should be taken into consideration by policy-makers. Finally, Canzian and Musolesi [114] analyze mobility patterns from GPS traces to answer whether mobile phones can be used to monitor individuals affected by depressive mood disorders. They develop a smartphone application that periodically collects the locations of the users and the answers to daily questionnaires that quantify their depressive mood. They find

Table 4 The table contains a subset of the information returned by a Twitter API

Id	Coordinates	Hashtags	Mentions	Text	Profile info	…
240556	null	#ny #dinner	[10214;452879]	…..	{…..}	…
4261063	NY	null	null	…..	{…..}	…
72096	42.10;10.2	#wellbeing	[964215]	…..	{…..}	…

If the user activates a localization system, the tweet also contains information on the position (longitude, latitude or city) from which the tweet is sent. Each tweet contains the information of the user profile and mentions or hashtags used in the text

a significant correlation between mobility trace characteristics and the depressive moods of individuals.

Some of these works using GPS data focus on exploring socioeconomic development (B3). Marchetti et al. [115] perform a study at regional level, analyzing GPS tracks from cars in Tuscany to extract measures of human mobility at province and municipality level. They find that there is a strong correlation between the mobility measures and a poverty index independently surveyed by the Italian official statistics institute. Smith et al. [116] use an automated fare collection data set of journeys made on the London rail system to build a classifier that identifies areas of the city with high economic deprivation. They highlight that, given its high precision, the classifier provides potential benefits for city planning and policy-making. Lathia et al. [117] use the same data set to find that more deprived areas tend to receive passenger flow from a higher number of other areas compared to less deprived areas, also uncovering some evidence of social segregation.

Another objective well-being dimension that is explored with GPS data is safety (B5). Robinson et al. [118] compare the spatial distribution of objective crime incidents and self-reported physical activity among adolescents in Massachusetts, between 2011 and 2012, and show that there is a positive association between them ($r = 0.72$, $p < 0.0001$). Ariel et al. [119] use GPS data to replicate findings published from US official research on the effect of hot spots policing for the prevention of crime in England and Wales and demonstrate that victim-generated crimes (the primary outcome measured in previous studies) increase in both the near vicinity and in catchment areas.

2.2.3 Social media data

Social media, such as Twitter, Facebook, and Instagram, can be considered as a digital database of information about online users, hence rendering individuals' online activities accessible for analysis. Given this enormous potential, researchers, governments, and corporations are turning their interest on social media to understand human behavior and interactions better [120]. Among all social media, Twitter is the most popular, since it provides public access to data through APIs with the least restrictive policy. The Twitter APIs return information about locations, date of the event, interactions with other users, or tags inserted in the tweet. Twitter also returns some information about the user profile. Table 4 illustrates an example of the structure of Twitter records.

Despite their indubitable usefulness, social media data may also encounter some concerns [121]. First of all, they may reflect social desirability biases, since individuals manage their online profiles [122]. Besides, social media users may not be as representative of the general population as traditional anonymized self-reports conducted through a chosen representative sample [123].

All dimensions of objective well-being are monitored through social media data, i.e., health (C1), job opportunities (C2), socioeconomic development (C3), environment (C4), safety (C5) and politics (C6).

Several studies provide valuable insights into how the analysis of social media data can lead to next-generation automated methodologies for public health (C1). As an example, Eichstaedt et al. [123] use Twitter data, in combination with atherosclerotic heart disease (AHD) mortality rates and country-level socioeconomic variables. They predict country-level heart disease mortality since the language expressed on Twitter reveals important psychological characteristics that are significantly associated with heart disease mortality risk. Besides, De Choudhury et al. [124] use Twitter data in combination with traditional depression screening test data for the detection and diagnose of the individuals' major depressive disorders and even to predict the likelihood of depression of individuals. Signorini et al. [125] use data from Twitter to track rapidly-evolving public sentiment concerning H1N1 and to measure actual disease activity. They show that Twitter can be used as a measure of public interest or concern about health-related events and that estimates of influenza-like illness derived from Twitter chatter accurately track reported disease levels. Paul et al. [126] incorporate in their forecasting models the historical influenza data and Twitter data. Lampos et al. [127] measure the prevalence of flu-like symptoms in the general UK population, based on the contents of Twitter, searching for symptom-related statements, turning this information into a flu-score and they obtain on average a statistically significant linear correlation which is higher than 95%. In a later work, the authors [128] instead of choosing the keywords and phrases themselves, they use machine learning algorithms to find out which words in the database of tweets occurred more often at times of elevated levels of flu, and they obtained very positive results. They claim that flu epidemics can be detected based on Twitter content. Chen and Yang [129] use individuals' tweets with spatiotemporally tagged information to demonstrate that people's healthy diet is elicited by exposure to their immediate food environment.

Regarding the monitoring of job opportunities (C2), Llorente et al. [130] quantify the extent to which deviations in diurnal rhythm, mobility patterns, and communication styles across regions relate to unemployment. For this purpose, they examine country-wide Twitter data describing 19 million geo-located messages and find that the regions exhibiting more diverse mobility fluxes, earlier diurnal rhythms, and more correct grammatical styles display lower unemployment rates. Antenucci et al. [131] use data from Twitter, from July 2011 to early November 2013, to create indexes of job loss, job search, and job posting. They derive signals by counting job-related phrases in tweets such as "lost my job". They construct social media indexes from the principal components of these signals and manage to track events that affect the job market in real-time, such as Hurricane Sandy and the federal government shutdown.

A large number of works in the literature focus on monitoring socioeconomic development from social media data (C3). Bollen et al. [132], in a further study, analyze data from Twitter and consider the emotions of traders, rather than their information gathering processes, suggesting that changes in the calmness of Twitter messages could be linked to changes in stock market prices. Still, regarding socioeconomic development, social media data are also extensively used to nowcast and forecast stock market prices and traded volumes. Seminal works in this field leverage information contained within investment discussion boards and blogs. For example, Bar-Haim et al. [133] use StockTwits data to uncover relevant correlations between Web-derived indicators and the stock market. In detail, they leverage sentiment scores of messages shared in the Yahoo message boards to find correlations with the stock market. In a different web platform study, De Choudhury et al. [134] try to find correlations between the stock market and blog communications. Last, Cresci et al. [135,136] assess the risks and vulnerability of stock markets to automation, manipulation, and disinformation, with the ultimate goal of safeguarding people's investments.

Researchers also use social media for the exploration of the environment dimension (C4). Avvenuti et al. [137] claim that the analysis of social media proves valuable for quickly acquiring situational awareness and estimates of the impact of disasters. As an example of the predictive power of social media, Kryvasheyeu et al. [138], Avvenuti et al. [139] and Mendoza et al. [140] demonstrate the viability of predicting or nowcasting the damage produced by earthquakes by analyzing social media communications in the aftermath of the event. The results of these models can also be displayed in real-time, interactive maps that highlight stricken areas and provide support to emergency responders. Notable examples of this kind are the systems developed by Avvenuti et al. [141,142]. Preis et al. [143] find that the number of photos taken and subsequently uploaded to Flickr with titles, descriptions, or tags related to Hurricane Sandy bears a striking correlation to the atmospheric pressure in the US state New Jersey. They claim that appropriate leverage of such information could be useful to policy-makers and emergency crisis managers.

Safety is another dimension that can be monitored using data from social media (C5). For example, Chen et al. [144] use Twitter data and create a model that predicts the specific time and location a crime occurs. This model combines kernel density estimation based on historical crime incidents and prediction via linear modeling with sentiment and weather predictors. By adding the latter determinants, they show that their model improves significantly with respect to existing models. Similarly, Boni et al. [145] use spatio-temporally tagged tweets and create a model for crime prediction. In particular, they combine real crime data with individuals' micro-level movement patterns extracted from Twitter and demonstrate improved predictions. Likewise, Kadar et al. [146] describe urban crime by using Foursquare and considering these data as a measurement for the ambient population of a neighborhood, to further describe crime levels. They also confirm that such models improve the traditional models, based on census data. Additionally, the city of Chicago applies text analytics on Twitter and 311 (the local emergency number) records to detect and prevent phenomena like rat infestations and to track civil unrest and violent crimes (CrimeScan and CityScan software) [147–149].

Finally, the politics dimension (C6) is extensively studied, in particular, during the last years with the rise of the political crisis across the world. Colleoni et al. [150] investigate the political homophily on Twitter to classify users as Democrats or as Republicans based on their tweets. They show that, in general, the former exhibit higher levels of political homophily than the latter. Goh et al. [151] use Facebook pages of a group of 12 politicians and demonstrate that political engagement can be achieved by creating social media consumption habits, as supported by the habit formation in consumption from macroeconomics. Similarly to the field of socioeconomic and financial analyses, social media data can be easily manipulated also for achieving political goals [152,153]. As such, results of political analyses based on social media should be carefully weighed to minimize issues related to biases and manipulations.

Table 5 The table shows an example of clinical records, including the pathology for which a patient is admitted to the hospital, the duration of hospitalization and the medicines she/he took

In date	Out date	Pathology	Medicines
01/02/2019	01/02/2019	Asthma	m1,m2,m3
03/02/2019	08/03/2019	Head trauma	m5

2.2.4 Health and fitness data

These data mainly consist of Electronic Health Records (EHRs) and mobile application data that are mainly used for monitoring the health dimension (D1). EHRs, initially created for the facilitation of the billing and patient care, are widely used for clinical studies and clinical risk prediction. Table 5 reports an example of clinical records concerning the hospitalization of some patients.

Out of a systematic review, Goldstein et al. [154] demonstrate both opportunities and challenges of EHRs. On the one hand, compared to the traditionally used cohort data developed and collected for research purposes (such as the Framingham Heart Study [155]), EHRs are cost-effective. In contrast with cohort data, EHRs can indeed be used for multiple health studies and, since they are collected at a high frequency, they allow a better prediction of near-term risk of events. On the other hand, EHRs include only individuals that have been ill or at least have had a clinic visit, which could generate a problem of representativeness. Moreover, they are not publicly available and might include limited time slots.

Researchers use EHRs to monitor several aspects of personal health (D1). For example, Sultana et al. [156] use the Integrated Primary Care Information (IPCI) database to look for elements that could contribute to traditional methodologies. For example, multimorbidity and polypharmacy are elements that could help in identifying frailty methodologies. They demonstrate that the Mini-Mental State Examination score, which is the most commonly recorded data item, could be potentially used as a frailty identifier. Ghaderighahfarokhi et al. [157] use medical records of newborns in the educational Hospitals affiliated to the Ilam University of Medical Sciences (from April 2015 to April 2016) to identify accurate predictors of Low Birth Weight (LBW). They demonstrate that LBW is a multi-factorial condition requiring a systematic and accurate program to be reduced, such as education through mass media, repeated monitoring of pregnancy, and others. Metzger et al. [158] use EHRs with Emergency Department patient visits in 2012, from Lyon University Hospital, to demonstrate that machine learning can contribute to more accurate estimations of suicide attempts in France, in relation to the current national surveillance system based on manual coding by emergency practitioners. Mhaskar et al. [159] investigate the 30 minutes prediction of blood glucose levels based on Continuous Glucose Monitoring (CGM). In particular, they use data from the DirecNet Central Laboratory, containing time series for 25 patients, who are less than 18 years old. By training a deep learning model on a data set designed to explore the performance of CGM devices in children with Type I diabetes, they demonstrate how deep neural networks can outperform shallow networks on this task. In addition, Santillana et al. [160] use a clinician's database, named as UpToDate, to predict influenza epidemics in the United States promptly. They show that digital disease surveillance tools based on experts' databases may be able to provide an alternative, reliable, and stable signal for accurate predictions of influenza outbreaks. Besides EHRs, mobile app data, such as lifestyle habits data concerning eating and physical activity behaviors, are used for the monitoring of objective well-being in terms of health (D1). These data demonstrate for once more that smartphones can contribute to research with valuable new insights, although they might apply biases towards people with lower socioeconomic status or towards people who are more interested in their health. In addition, such data collected through web surveys for research purposes might bring the disadvantages discussed before. A critical study using mobile app data is conducted by Althoff et al.[161]. They use a data set consisted of physical activity for 717,527 Apple iPhone smartphone users of the Azumio Argus app, which tracks users' diet and fitness and other healthy behaviors, between July 2013 and December 2014. They demonstrate inequality in how the activity is distributed within countries and that this inequality is a better predictor of obesity than average activity level. Similarly, Hayeri [162] uses continuous glucose monitors (CGM) and fitness wearables (Fitbit) to predict blood glucose values. The study uses data gathered from each participant for 60 days, where the data from the first 30 days are used to train the algorithm and the remaining 30 days to test the predictions. On average, the software is able to predict a user's future glucose values with a 93% accuracy rate for 60-mins ahead of time.

2.2.5 News

News data sources, such as the GDELT database [163], contain information extracted from the news of newspapers around the world. News records generally describe a variety of subject domains (e.g., economic events, political events), represent a wide range of targets (e.g., opposing politicians) [164] and are continuously updated, containing even archived historical news of the last decades. Nevertheless, such data contain three main biases [165]: the gatekeeping bias, i.e., the editors or the journalists decide on which event to publish; the coverage bias, related to the coverage of an event (e.g., western countries are over-covered, whereas African countries are under-covered); the statement bias, when the

Table 6 Subset of the main fields provided by GDELT platform

EventCode	EventCategory	EventTone	Date	Country code	Url
815176338	Arrest, detain	− 70	20180110	US	http://tiny.cc/s5s16y
815176339	Use conventional military force	− 30	20180110	UK	…
815176340	Consider policy option	+ 25	20180110	IT	…

content written by the journalist, even if tried to be objective, is favorable or unfavorable towards certain events. Table 6 shows an example of news records.

News records are used to measure health (E1), socioeconomic development (E3), environment (E4), and politics (E6) dimensions of objective well-being.

Emerging infectious diseases and the rise of modern technology have generated new demands and possibilities for disease surveillance and response (E1). Growing numbers of outbreak reports must be assessed rapidly so that control efforts can be initiated. For example, the World Health Organization (WHO) sets up a process for timely disease outbreak verification to convert large amounts of data from some 600 sources, including all major news wires, newspapers, and biomedical journals, into accurate information for suitable action [166,167]. Brownstein et al. [168] in a similar effort, create HealthMap, a freely accessible, automated real-time system that monitors, organizes, integrates, filters, visualizes, and disseminates online information about emerging diseases. Wilson et al. [169] use the HealthMap project to monitor listeriosis. Chunara et al. [170] use social and news media to validly estimate the 2010 Haitian cholera outbreak.

News records on financial affairs and financial markets are intrinsically interlinked (E3). Alanyali et al. [171] quantify the relation between movements in financial news and movements in financial markets by exploiting a corpus of six years of financial news from 2007 to 2012 from the Financial Times. Their results suggest that greater interest in a company in the news is related to greater interest in the corresponding company in stock markets. Lillo et al. [172] show that the flux of news of the previous day affects the trading activity of companies, households, and foreign investors and the dynamics of volatility.

News can also help capturing the environmental dimension of well-being (E4). As an example, Kleinschmit et al. [173] investigate 394 articles on forest and climate change published in the Swedish newspaper Dagens Nyheter from 1992 to 2009. They show that there has been an increasing discussion on forests in a changing climate over the last 18 years from both scientists and politicians. The increased number of these news events correlate with real environmental events happening internationally. Similarly, Boykoff [174] uses data extracted from the Vanderbilt University Television News Archive, consisting of television news from US news broadcasts (e.g., ABC World News Tonight) for the period between 1995 and 2004. He demonstrates that 70% of the US television news provide balanced coverage on anthropogenic contributions to climate change compared to natural radiative forcing. He also shows that there is a significant difference between this television coverage and scientific consensus on the topic.

News records are also used to understand the coverage of political issues (E6). Van Aelst and De Swert [175] use daily news of politics of campaign periods, extracted from the Electronical News Archive over the 2003 to 2006 period, and show that campaign periods have a high impact on the amount, style and actors of the political news in Belgium. To the best of our knowledge, the dimension politics (E6) has not yet been adequately explored through news data and constitutes inspiration for future research.

2.2.6 Scanner data

Scanner data are generated by point-of-sales terminals in shops and provide information at the level of the single product. Sales terminals record each transaction, and the resultant data can provide considerable insights into consumer purchasing patterns. They can be obtained from a wide variety of retailers: supermarkets, pharmacies, do-it-yourself stores, home electronics or clothing shops, and many others [176].

Scanner data are used from social researchers, as they can offer useful detailed information and the possibility to model the dynamic behavior of households, as well as to control for unobservable time-invariant characteristics [177]. Also, scanner data provide information over long periods of time than only one day or a couple of weeks. This happens because the final data used are produced from customers that purchase several items on each store visit, for several store visits, over a period of time [178,179]. It is also worth mentioning that scanner data can contribute to the improvement of the quality of the Harmonized Index of Consumer Prices (HICP) [180]. However, using scanner data is challenging since researchers are dependent on the retailer’s permission, and they should also overcome the legal constraints in order to obtain them [179]. Table 7 shows an example of supermarket records.

Scanner data are used to measure health (F1), socioeconomic development (F3), and environment (F4) dimensions of objective well-being.

Table 7 Subset of the main fields provided by a supermarket database for the purchases in different shops

Id	Customer	Timestamp	Place	Receipt	Items
2018020156287	109745368	2018-02-01 17:30:14	Pisa, Italy	2018020101567	[bread, milk, eggs, tissues]
2018020578256	104827423	2018-02-05 10:14:57	Torino, Italy	2018020500234	…
2018020743624	012753862	2018-02-07 19:57:00	Florence, Italy	2018020721987	…

To begin with, researchers use scanner data to monitor several aspects of public health (F1). For example, pharmaceutical sales may be used to predict changes in clinical conditions with a useful time lead. Magruder et al. [181] find a 90% correlation between flu-related drug sales and physician diagnoses of acute respiratory conditions, at several subregions of the National Capital Area. They show that these sales occur approximately three days before the physician-patient encounter. Scanner data are also used to study the nutrients and saturated fat of several food categories and their implications on personal health. For example, Griffith et al. [177] use supermarket scanner data from the UK to study the nutrients in foods. They show that there is a lot of variation in nutrients at individual product level, even with food categories such as butter, which are very narrow. Bonnet et al. [182] use data from French supermarkets to explore consumer behavior with respect to the consumption of saturated fat, while Griffith et al.[183] model the potential impact of a tax on saturated fats. Finally, Janssen et al. [184] use scanner data from the Nielsen Consumer Panel data set that covers the years from 2004 to 2017. They aim to identify households with a pregnant household member and also to estimate the effect during and after pregnancy on alcohol purchases and relative expenditure on fruit and vegetables. Results show that during and after pregnancy, households reduce their alcohol purchases by 22–27%. In contrast, the relative expenditure on fruit and vegetables does not increase during pregnancy but decreases post-pregnancy by 19%.

The majority of studies with scanner data focus on exploring the socioeconomic development (F3). Van der et al. [186] introduce a new method for computing the Dutch Consumer Price Index (CPI) based on supermarket scanner data. In the meanwhile, in 2017, Eurostat issued a practical guide for processing supermarket scanner data to calculate the CPIs of EU countries in order to ensure the comparability of the values across Europe, as well as to modernize the official statistics [179]. Silver et al. [187] outline the potential use of scanner data from retailers for the measurement of inflation. They use monthly scanner data for television sets in 1998 in the UK to study the two primary forms of bias in CPIs. Moreover, Pennacchioli et al. [188] study the retail activity of the customer subset of an Italian supermarket chain. They discover that highly ranked customers, with more sophisticated needs, tend to buy niche products, i.e., low-ranked products. On the other hand, low-ranked, low purchase volume customers tend to buy only high-ranked products, very popular products that everyone buys. In addition, Sobolevsky et al. [189] use a complete set of bank card transactions in 2011 in Spain and demonstrate that there is a clear correlation between individual spending behavior and official socioeconomic indexes denoting the quality of life.

Finally, researchers use scanner data to monitor the impact of humans on the environment (F4). Panzone et al. [190] use scanner data from the largest UK food retailer for the creation of an Environmentally Sensitive Shopper (ESS) index measuring the environmental sustainability of food consumption at household level. In addition, Gadema et al. [191] use data from UK supermarket shoppers to examine whether carbon footprinting and labeling food products are tools that could facilitate consumers to make greener purchasing decisions. They claim that this could be a sensible way to potentially achieve a low carbon future. Food waste is a significant problem in modern society and carries considerable social, economic, and environmental costs. For example, Brancoli et al. [192] use scanner data to analyze the impacts of food waste at a supermarket in Sweden. They discover the importance of not only measuring food waste in terms of mass but also in terms of environmental impacts and economic costs. They also show that meat and bread waste contribute the most to the environmental footprint of the supermarket. Last, Scholz et al. [193] analyze food waste data of six Swedish supermarkets from 2010 to 2012 in terms of mass and carbon footprint. They calculate the wastage carbon footprint for fresh products such as meat, deli, cheese, dairy, and fruits and vegetables.

2.2.7 Web search queries

Web search queries data report the frequency of specific terms over time, entered into a web search engine from users to satisfy their information needs. Data are represented as time series of the frequency, and therefore we do not provide an example of search queries records in this paper.

Comparing to other data sources that require customized and often complicated collection strategies, search data can be collected for many domains simultaneously. They can also be easily analyzed across several countries or regions in real-time. Search data are often helpful in making forecasts. However, their utility for predicting real-world events is based on convenience, speed, and flexibility and has less to do with their superiority over other data sources. Goel et al. [194] provide a useful survey in this area and describe some of the limitations of this data source. First, for different domains, the size of the relevant population varies considerably, along with difficulty in identifying relevant queries. Additionally, in specific domains, searching may be more closely tied to the measured outcomes than in others.

Web search queries data are used to measure health (G1), job opportunities (G2), socioeconomic development (G3), safety (G5), and politics (G6) dimensions of objective well-being.

Public health is a dimension of well-being that is explored through web search queries (G1). In order to improve early detection, researchers monitor health-seeking behavior in the form of web search queries, which are submitted by millions of users around the world every day. For example, Cooper et al. [195] study Yahoo! search activity related to cancer in the USA. They find out that the Yahoo! search activity associated with cancer correlates with the estimated cancer incidence and estimated cancer mortality. Polgreen et al. [196] show that search volume for handpicked influenza-related queries is correlated with the reported number of cases over the period from 2004 to 2008. Hulth et al. [197] find similar results in a study of search queries submitted on a Swedish medical Web site. Yuan et al.[198] monitor influenza epidemics in China with search queries from Baidu. Additionally, an automated procedure for identifying informative queries is described by Ginsberg et al. [199]. Based on that, Google Flu Trends [200] was introduced by Google in 2008 to provide real-time estimates of flu incidence for more than 25 countries and to help predict outbreaks of flu. Nsoesie et al. [201] present a framework for near real-time forecast of influenza epidemics using web-based estimates of influenza activity from Google Flu Trends for 2004–2005, 2007–2008, and 2012–2013 flu seasons. Yang et al. [202] use Google Flu Trends and historical data to infer the evolving epidemiological features of influenza and its impacts among the large population during 2003–2013, including the 2009 pandemic. Wilson et al. [203] use data from Google Flu Trends to study the spread of the pandemic H1N1 influenza in New Zealand during 2009. Furthermore, Chan and Althouse [204,205] use Google queries to monitor Dengue epidemics, Dukic et al. [206] to predict hospitalizations for methicillin-resistant Staphylococcus aureus infections and Ocampo et al. [207] for malaria surveillance. Moreover, Yang et al. [208] evaluate the association between suicide and Google searches trends for 37 suicide-related terms representing major known risks of suicide in Taipei City, Taiwan, from 2004 to 2009. Their results show that a set of suicide-related search terms, the trends of which either temporally coincided or preceded trends of suicide data, are associated with suicide death. Searches for "major depression" and "divorce", for example, account for at most, 30.2% of the variance in suicide data. McCarthy [209] uses annually-averaged Google search activity for "suicide" from the same period, from 2004 to 2009 to study suicide rate data in the United States. The study shows that searches for most medical, familial, and socioeconomic terms precede suicide deaths, and most searches for psychiatric-related terms coincide with suicide data. In a later work, Kristoufek et al. [210], use Google data from 2004 to 2013 in combination with suicide occurrences data to estimate the number of suicide occurrences in England. Finally, Adler et al. [211] combine official statistics on demographic information with data generated through search queries from Bing, between November 2016 and February 2017, to gain insight into suicide rates per state in India. In this way, their search data work as a proxy for unmeasured (hidden) factors corresponding to suicide rates.

The first to explore the job opportunities dimension (G2), are Ettredge et al. [212] as they find that counts of the top 300 search terms during from 2001 to 2003 are correlated with US Bureau of Labor Statistics unemployment figures. Later on, Askitas et al. [213], D'Amuri et al. [214], Suhoy et al. [215] confirm the value of search data in forecasting unemployment in the US, Germany, and Israel. Baker et al. [216] use Google search data to examine how job search responds to extensions of unemployment payments. Finally, McLaren et al. [217] summarise how online search data can be used for economic nowcasting by central banks. They show that the volume of online searches can be used as indicators of economic activity, more specifically for unemployment and housing markets in the United Kingdom.

Researchers use search queries to monitor socioeconomic development (G3) as well. Choi and Varian [218,219] consider Google Trends as a source of data on real-time economic activity, and they show that by using its query indices accurate predictions can, for example, be made for retail, automotive, etc., and could be helpful for short-term economic prediction or nowcasting. Koop and Onorante [220] use Dynamic Model Selection (DMS) methods, which allow for model switching between time-varying parameter regression models. They extend the DMS methodology by allowing Google variables to determine the nowcasting model to be used at each point in time. Guzman [221] examines Google data as a predictor of inflation. Additionally, Preis et al. [222] provide evidence that search engine query data and US stock market fluctuations are correlated. In a later [223] work, they analyze changes in Google query volumes for search terms related to finance, and they find patterns that may be interpreted as "early warning signs" of stock market moves. Furthermore, Curme et al. [224] present a method that allows identifying topics for which levels of online interest change before large movements of the Standard & Poor's 500 index (S&P 500). They find that search volumes from Google

Table 8 Example of the information provided by users of influenzanet

User	Age	Gender	Date	Highest temperature	Symptoms
784590	35	M	2017-12-03	38.0°	[cough, sore throat]
275173	28	F	2018-01-05	36.6°	[no symptoms]
428415	64	M	2018-04-13	38.2°	[tired, runny nose]

related to politics and business can be linked to subsequent stock market moves. This demonstration of a connection between stock market transaction volume and search volume is also replicated using Yahoo! data, where Bordino et al. [225] show that query volumes precede in many cases peaks of trading by one day or more. Finally, Moat et al. [226] show that data on views of Wikipedia pages can also be related to market movements, providing evidence that increases in the number of views of financially related pages on Wikipedia can be detected before stock market falls.

Search data are also used for the exploration of safety (G5). Qi et al. [227] show that a simple low-level indicator of civil unrest can be obtained from online data at an aggregate level through Google Trends or similar tools. The study covers countries across Latin America from 2011 to 2014 in which diverse civil unrest events took place. In each case, they find that the combination of the volume and momentum of searches from Google Trends surrounding pairs of simple keywords, tailored for the specific cultural setting, provide useful indicators of periods of civil unrest. Qi et al. [228] study online search activity from Google Trends surrounding the topics of social unrest over several countries in Latin America from 2011 to 2014. They find that the volume and momentum of searches surrounding mass protest language, can detect—and may even pre-empt—the macroscopic on-street activity. They also find that the most crucial search keywords differ subtlety from country to country, even though the language may be the same. They explain this by the fact that civil unrest is a time-varying coordinated interaction between individuals, groups, or populations within a given cultural and socioeconomic setting.

Finally, the politics dimension is explored with search data (G6). Chykina et al. [229] study how Google Trends can be used to examine issue salience for hard-to-survey mass populations in the US, from 2010 to 2017. They apply this method to immigrant concerns over deportation. They show that anxieties over removal increase in response to (potential) policy changes, such as immigration policies that are considered in the wake of Donald Trump's election. Reilly et al. [230] use Google search activity for ballot measures' names and topics in a state one week before the 2008 Presidential election, and they find that they correlate with actual participation on those ballot measures. Their result demonstrates that the more Internet searches there are for a ballot measure, the less likely voters are to roll-off (not answering the question) and establish the validity for this data for a critical topic in state politics research.

2.2.8 Crowdsourced data

Kleemann and Rieder [231], in 2008, have defined crowdsourcing as the "the intentional mobilization for commercial exploitation of creative ideas and other forms of work performed by consumers". In other words, crowdsourcing involves obtaining work, information, or opinions from a large group of people who submit their data via the Internet, smartphone apps, etc. Naturally, crowdsourcing brings several advantages. Crowdsourcing can provide researchers with a huge amount of data, which can be accessed quickly and at a relatively low cost. Besides, comparing to traditional research (such as studies using traditional surveys), the use of crowdsourcing can provide researchers with data from samples that are more diverse [232]. However, crowdsourcing yields various challenges, as well. Firstly, crowdsourcing may bring relatively low-quality results, e.g., a participant of a crowdsourced study may intentionally give wrong answers. Secondly, mobile platforms pose new challenges for crowdsourced data management. Table 8 shows an example of crowdsourced data.

Crowdsourced data are used to capture all dimensions of objective well-being, i.e., health (H1), job opportunities (H2), socioeconomic development (H3), environment (H4), safety (H5) and politics (H6) dimensions of well-being.

To improve early detection, researchers started monitoring the health of individuals (H1) through crowdsourced self-reporting mobile apps, such as Influenzanet (Europe) [233], Flutracking (Australia) [234], and Flu Near You (United States) [235]. Hashemian et al. [236] introduce iEpi, an end-to-end system for epidemiologists and public health workers to collect, visualize, and analyze contextual microdata through smartphones. Additionally, Madan et al. [237] use data from a smartphone application provided to university students to study their health state. Participants fill out self-report surveys related to their health habits, diet, exercise, weight changes, daily symptoms related to common colds, fever, influenza, and mental health. The researchers find that phone-based features can be used to predict changes in health, such as common colds, influenza, and stress. For longer-term health outcomes such as obesity, they find that weight changes of participants are correlated with exposure to peers who gain weight in the same period. Finally, Martinucci et al. [238] study Gastroesophageal Reflux Disease (GERD) symptoms among Italian university students from a data set collected from a web-app. The app allows users a self–diagnosis for the gastrointestinal disturbances through a simple questionnaire and data about the students' food consumption at the university canteen. They show that 792 students reported typical GERD symptoms to occur at least weekly. Among all users, females, smokers, and high in BMI students tend to show increased GERD values.

Researchers use crowdsourced data to explore the job opportunities dimension (H2) and the direct socioeconomic benefits associated with it. For example, Green et al. [239] use the crowdsourced employer review website named Glassdoor, an online crowdsourced employer branding platform, to explore employees' satisfaction and work–life balance. This exploration is preliminary for the direct economic benefit and most important finding of the study; companies experiencing improvements in employer ratings are significantly associated with future stock returns, comparing to companies with declines in employer rating. Similarly, Dabirian et al. [240] analyze reviews of the highest and lowest-ranked employers on Glassdoor. Using IBM Watson to analyze the data, they show how employers could use crowdsourced employer branding intelligence to turn into a workplace that attracts highly qualified employees. Furthermore, Könsgen et al. [241] analyze employee reviews data, listed on the German employee review site named Kununu.de, combined with 2×2×2 between-subjects experimental design. Results show that such studies can complement the research on the online reputation by underlying the relevance of discrepant reviews for job candidates' application intentions.

Crowdsourced data are also used to estimate the socioeconomic (H3) well-being. For example, Tingzon et al. [242] show the feasibility to map poverty by combing crowdsourced geospatial information with nighttime lights, daytime satellite imagery, and human settlement data. In particular, they use the popular geospatial data crowd-sourcing platform named OpenStreetMap [243] to map poverty in the Philippines. Similarly, Piaggesi et al. [244] use OpenStreetMap [243] crowdsourced data merged with official data at a city scale. They demonstrate the possibility of estimating the socioeconomic conditions of different neighborhoods of five different cities in North and South America. In order to increase the efficiency of direct money transfers to impoverished villages in Kenya and Uganda, Abelson et al. [245] develop and deploy a crowdsourcing interface to obtain labeled satellite imagery training data. They train and deploy a predictive model for detecting impoverished villages. Their estimations are leveraged to build a fine-scale heat map of poverty that is used to recommend donations to the most impoverished villages.

Crowdsourcing is also used to capture the environmental dimension of well-being (H4). There are plenty of examples of crowdsourcing platforms for emergency management, such as Ushahidi [246], where volunteers provide updated environmental information in the aftermath of mass emergencies. These platforms are shown to contribute significantly to organizing a prompt emergency response [247]. Another category of crowdsourced platforms is the so-called citizens' observatories [248], a community-based network of environmental monitoring and information systems. On these platforms volunteers monitor and provide data about a plethora of environmental dimensions, such as comprising water availability and water quality, air pollution, land use, and flood risk management [249]. As an example, Schneider et al. [250] combine crowdsourced data from the EU-funded CITI-SENSE project, which measures the air-quality with data obtained from statistical or deterministic air quality models. Their goal is to present a novel data fusion-based technique for combining real-time crowdsourced observations with model output that maps the urban air quality in detail. This could help users find the least polluted routes or control their exposure to pollution while moving around the city. Besides, Meier et al. [251] use crowdsourced atmospheric data from Netatmo weather stations in the city of Berlin, as well as available metadata to explore the urban atmosphere. Results show a distinctive urban heat island pattern in Berlin during the night and are also validated, confirming that crowdsourced atmospheric data can contribute to advancement in climate research. Similarly, Chapman et al. [252] use Netatmo weather station crowdsourced data to quantify the urban heat island in the city of London over the summer of 2015. Their results are similar to previous studies with official data and are therefore validated.

Crowdsourced data are considered an important data source for studying safety (H5). Suzanne Goodney et al. [253] map violence against women with the use of a crowdsourced app named as Safecity.in, which includes anonymous reporting of violence against women. The goal of the study is to highlight the importance of crowd mapping violence, as it can make women aware of potentially dangerous locales, encourage violence reporting, and provide advice on practical solutions for navigating street harassment and assault in public buses. Furthermore, Gosselt et al. [254] use the Internet Movie Database (www.imdb.com) to study the violent behavior and victimization of male and female film characters over time in the United States. In particular, using IMDb synopsis texts, they analyze reviewers' movie descriptions. They demonstrate that both perpetrators and victims are mainly male, as well as that violence becomes less severe and more often non-deadly over the years. Researchers underline the future potentiality of using such data sources to explore matching results with actual crime figures. Additionally, Ozkan et al. [255] use crowdsourced police-involved killings data from FatalEncounters.org, as well as media data, to control whether police killings is counted and reported correctly in the aforementioned unofficial data, as compared to official data in the city of Dallas. Results mostly show consistency between all data sources. In conjunction with social media and crowdsourcing data sources, as well as environmental and safety dimensions, Avvenuti et al. [256] collect targeted and detailed information from people involved in natural disasters through crowdsourcing surveys via social media. These data are used to monitor unfolding disasters better and to monitor their consequences (i.e., damage caused)

Last, crowdsourced data are also used to study the politics dimension (H6) of objective well-being. For instance, crowdsourced data have been used within NGOs to set strategic priorities and involvement in the referendum activities based on participants' responses to a survey [257]. Yasseri and Bright [258] use Wikipedia traffic data for electoral prediction. In particular, they get insights about changes in overall turnout at elections and changes in vote share for certain parties. Furthermore, Gellers [259] explores whether crowdsourcing can overcome the democratic deficit in global environmental governance. He uses data from the United Nations MY World survey, a multi-year (2012–2015) global poll designed to identify post-2015 development priorities, as well as e-discussions data, organized by the UNDG and the thematic consultation on environmental sustainability ran from November 2012 to July 2013. Results suggest that although crowdsourcing may present an attractive technological approach to enhance participation in global governance, ultimately, the representativeness of this participation and the legitimacy of the policy results depend on the way the contributions are sought and filtered by international organizations.

3 Measuring subjective well-being

"Subjective well-being", the scientific term of happiness, is a central value in people's lives, and reflections for its definition have arisen ever since antiquity. Aristotle has expressed his interest on the topic claiming that human well-being, labeled as eudaimonia ($\varepsilon\upsilon\delta\alpha\iota\mu o\nu\acute{\iota}\alpha$: Eu=Good, Daimon=spirit), is an activity of the soul expressing complete virtue [260]. During the last decades, researchers have focused on identifying the critical dimensions and the relevant determinants that can positively or negatively affect human well-being, hence providing a perspective different from the philosophical definition that Aristotle has been contemplating about. Since humans are conscious beings, they can subjectively evaluate their appreciation of life, labeled "subjective well-being" or happiness. In particular, happiness can be defined as satisfaction with life in general, or as sociologist Veenhoven (1984) suggests, as the degree to which an individual judges the overall quality of her life-as-a-whole favorably. Similarly, psychologist Diener [261] defines happiness as people's affective and cognitive evaluations of life. Veenhoven [262] shows that people use two sources of information to evaluate their appreciation with life-as-a-whole: affects and thoughts. The first source of information captures people's feelings, emotions, and moods, the so-called hedonic level of affect (or simply called emotional component). In particular, he underlines that to avoid neglecting crucial information about precedent and subsequent events, researchers should separate between positive and negative affects. On the other hand, the second source of information is the contentment component (or simply called structural component), concerning people's thoughts and capturing whether people's life expectations have been fulfilled, according to their cultural or societal standards, and lead them to evaluate their life satisfaction. These two components, the hedonic level of affect and the contentment component, determine the overall happiness.

This concept of happiness, compared to the traditional macroeconomic measurements, such as GDP, inflation and national income (see, e.g., Alesina et al. [263]) can capture the variations of people's perceived well-being [11,12]. It is also worth mentioning the controversy surrounding the relationship between national income and national happiness, identified by Easterlin [30]. According to the Easterlin paradox, temporary changes in income both within and between nations directly affect happiness, but over time happiness does not trend upward as income continues to grow.

Considering its subjective nature, researchers frequently measure happiness by self-report rating scales. Nevertheless, the most widely used are global reports, using the single-item scale, such as the Positive And Negative Affect Scale (PANAS) [264,265]. Self-report measures are reliable since they provide accuracy and temporal stability, they are valid for community surveys and cross-cultural comparisons, and they can capture happiness as life-as-a-whole, as well as domain satisfactions [266–269]. Examples of self-reported surveys are the Gallup World Poll (e.g., study by Deaton [270]) and the World Values survey (e.g., study by Easterlin et al. [271]), which capture the worldwide happiness; the Gallup-Healthways Well-being index (e.g., study by Kahneman and Deaton [272]), the British Household Panel Survey (e.g., study by Frijters et al. [273]) and the Eurobarometer (e.g., study by Stevenson et al. [274]), which capture the happiness at local level. Although self-report surveys are widely used for the measurement of happiness, some factors might influence the results. For example, the type of questions asked before the happiness questions, as well as the individuals' mood at the time of the well-being rating, might disturb the results. Deaton and Stone [275] demonstrate a high item-order effect because of political questions coming before happiness questions. Also, substantial current-mood effects on happiness judgments are generated because of weather conditions, since they affect people's thoughts, feel-

Table 9 Pros and cons for each traditional data source and new data source used for the measurement of subjective well-being

Data source	Pros	Cons
Surveys - traditional data source	Accurate, temporal stability, valid for community surveys and cross-cultural comparisons, valid for capturing happiness as-a-whole and satisfaction domains	Item-order effect bias, current-mood effects, neglected temporal resolution
Ecological Momentary Assessment (EMA) - traditional data source	Measurement of the affective component, reduced retrospective biases, measurement of moment-to-moment variation of emotions	Disturbance of normal activities
Day Reconstruction Method (DRM)- traditional data source	Measurement of the affective component, time-budget information, reduced respondent burden	Neglected moment-to-moment variation of emotions
Social Media (Twitter, etc)-new data source	Continuously updated user-generated content, elimination of social desirability effect, few barriers in data extraction (Twitter)	Social desirability biases, non-population representative
Google Trends-new data source	Timeliness, observation of people's behavior	Interpretability of the value of the series, comparability of time series of different terms on a given day
Crowdsourcing-new data source	Measurement of daily behavior and activity	Use of self-reports, paid participation of users
News-new data source	Variety of data (e.g., text data), variety of subject domains, range of targets, archived historical news	Gatekeeping bias, coverage bias, statement bias

ings, and behavior [276,277]. Finally, because global reports are abstract general ratings of happiness over a long period, they neglect temporal resolution.

Diversely, researchers use Ecological Momentary Assessment (EMA) and Day Reconstruction Method (DRM) that are momentary diary self-report measures of happiness. They are designed to capture the affective components of happiness and reduce recall biases and heuristics [269]. In particular, EMA is a longitudinal research methodology that asks participants to report their feelings, thoughts, and emotions at the moment or right after each of their activities, avoiding retrospective biases and maximizing the accuracy of the assessments [278]. Similarly, DRM asks participants to reconstruct their daily life activities systemically and their experiences of the preceding days. It does not capture the moment-to-moment variation of emotions, as EMA does, but it avoids disturbing normal activities, requires less respondent burden [279] and captures time-budget information more efficiently [280]. Shiffman et al. [281] show that global reports of happiness are more predictive of future behaviors than momentary methodologies. Therefore, taking into consideration the pros and cons discussed above, researchers suggest a multi-method assessment, combing both global and momentary methods, to reach valid and accurate results [269,282].

The first rows of Table 9 provide a summary of the traditional data sources, as well as their pros and their cons, as discussed previously. The remaining rows are explained later.

3.1 The dimensions of subjective well-being

Over the years, researchers studied subjective well-being and have identified the dimensions and the relevant determinants that can positively or negatively affect human well-being. Some studies rely on small data sets (e.g., review by Diener and Seligman[283]) reflecting the psychologists' interest, such as personality, and some others use larger data sets, such as panel data (e.g., review by Dolan et al. [29]) reflecting the economists' interest. These studies, conducted with the use of traditional data sources, and in particular with surveys, have shed more light on identifying in detail the determinants of happiness, which we divide into five main dimensions explained below:

3.1.1 Human genes

Evidence shows that one of the most important predictors of happiness is human genes, which is fairly heritable, with 30% to 50% range, since there is a variation on the results across studies [13–20]. Therefore, on average, about 40% of the variance of individual differences in happiness scores is accounted for by genes. Personality, which falls under our genetic makeup, can distinguish between happy and unhappy personalities. For example, extraverted individuals are happier to anxious and worried ones [284]. People higher in self-esteem are less likely to suffer from depression [29]. In addition, studies undertaken with data across different countries and periods of time, find influences of the following results: age has a U-shaped effect on happiness, with the highest level of happiness on the youngest and the oldest age and the lowest level of happiness on the middle age, between 32 and 50 years [29]; women are either happier than men, or there is no significant difference between them in almost all 73 countries investigated [285]. However, these results should be carefully interpreted. For example, Deaton and Tortora [286] show that the U-shaped relationship between happiness and age in West countries turns into a linear relationship in sub-Saharan countries, where there is unavailability of social services for older people.

3.1.2 Universal needs

According to the evolutionary theory [22] and human's inherent growth tendencies [23], basic and psychological needs play an important role on happiness and are considered to be universal. In fact, Tay et al. [24], in a research conducted across 123 countries, show that life evaluation is associated with having basic and psychological needs, such as food and shelter, met ($r = 0.31$)[1]; positive affects are associated with the fulfillment of social needs ($r = 0.29$)[1] and the respect gained from other people ($r = 0.36$)[1]; negative affects are associated with the fulfillment of basic needs ($r = -0.17$)[1], respect gained from others ($r = -0.20$)[1] and autonomy needs ($r = -0.18$)[1] in terms of the degree of freedom in life. Therefore, according to Veenhoven and Ehrhardt's Livability theory [287], some societies have a better quality of life because they highly satisfy the aforementioned universal needs. It should be noted that each of these basic and psychological needs is independent of one another, meaning that each of them is influencing happiness beyond the effects of others.

3.1.3 Social environment

Many determinants fall under this dimension and can explain changes in the reported level of happiness. To begin with, education is an important determinant, which needs to be carefully studied since there is controversial evidence of its effects on happiness. Some studies of happiness economics suggest an insignificant relationship between higher education and happiness, whereas some others show a negative relationship between them [25–28]. On the other hand, other studies show that educated individuals tend to report more positive emotions and less negative ones, as well as more satisfaction with most domains of their life, such as financial, employment opportunities, etc., even when controlling for non-economic factors, such as marriage [288,289]. Besides, studies show that health is an important determinant, with psychological health to be more strongly correlated to happiness, than physical health (e.g., review by Dolan et al. [29]). Climate is another determinant, which appears to have effects on happiness. Rehdanz and Maddison's [290] study gives a reasonable indication that extreme weather is damaging to happiness. Moreover, living in an urban or rural area seems to influence happiness. In particular, living in big cities negatively affects happiness, whereas living in rural areas positively affects it (e.g., Hudson and Kyklos [291] for Europe; Hayo [292] for Eastern Europe). On the contrary, Rehdanz and Maddison [290] show different results on urbanization and ruralization. They demonstrate that population density does not affect happiness. Another important determinant is exercising. Naturally, Ferrer-i-Carbonell and Gowdy [293] show that people who exercise tend to have higher levels of happiness.

3.1.4 Economic environment

Income is one of the most discussed economic determinants of happiness. Easterlin [30] argues that while happiness and income show a positive relationship within nations, they show weak or no association between nations. He also shows that, across countries, although a relationship between happiness and income holds in the short-run, this is not the case over time. However, time-series and panel analyses across countries show that there is a positive relationship between income and happiness also in the long-run [32–34]. Veenhoven [31] challenges Easterlin's findings by arguing that people's happiness highly depends on the satisfaction of basic and psychological needs covered by income, which is more an absolute standard, than a relative standard. To the present day, studies support both arguments. Another source contributing to this debate is individual-level income or wealth and happiness data studies. For example, a longitudinal study with a sample of about 33,000 individuals shows that after two years, lottery winners rated higher happiness than non-lottery winners [294]. These contradicting findings confirm the complexity of the interpretation of the role of income and wealth on happiness, since the potential positive relationship between them may be moderated by other factors. For example, in the case of natural disasters, wealthier countries are more economically capable of providing financial aid to the people affected by the event [295]. Employment falls under the economic dimension category (see, e.g., [296,297]), likewise income. Evidence shows that unemployed individuals report lower happiness than employed individuals [298]. In particular, Knabe et al. [299] demonstrate that unemployment has a substantial relationship with diminished cognitive well-being, but does not decrease affective well-being.

[1] Zero-Order Correlations of needs and subjective well-being for the world.

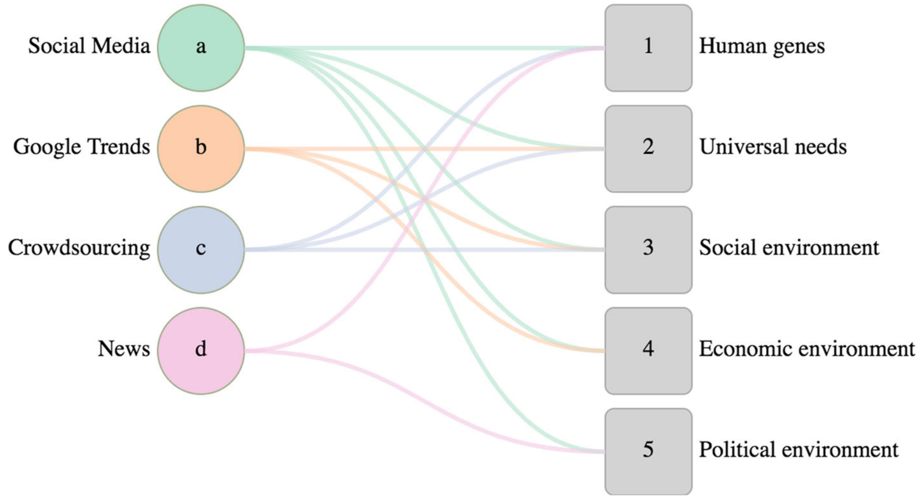

Fig. 2 The figure relates the sources of data (left) with the dimensions of the subjective well-being (right)

3.1.5 Political environment

As discussed previously, there are also political determinants associated with happiness. For example, Radcliff et al. [35] examine the effect of direct democracy, and in particular, the effect of the use of initiatives on happiness. They show that an individual's happiness is higher in states where not only initiatives are permitted, but also policy-makers depend on these initiatives to form the political system. Political freedom falls under this dimension, as well. Veenhoven [36] shows that political freedom is highly correlated with happiness in developed countries. Another political determinant associated with happiness is social hierarchy in terms of the differences in power and prestige. Brule and Veenhoven [300] show that in northern and southern European countries, people are less happy in hierarchical societies. Last, social trust [301] and government quality [302] are political determinants that are substantially associated with happiness.

3.2 Data sources for monitoring the dimensions of subjective well-being

Similarly to Fig. 1 on objective well-being, Fig. 2 describes the new data sources (left) that have been used to estimate one or more dimensions of subjective well-being (right). The presence of a link in Fig. 2 between a data source and a dimension indicates that there are papers in the literature on monitoring that dimension with that data source. For example, b4 indicates the link between Google Trends data (b) and economic environment (4).

In this section, we describe, for each data source, its features (e.g., the process of data collection, its biases and limitations) and the main works in the literature that use it to measure several dimensions of subjective well-being. Table 9 provides a summary of the new data sources used to explore happiness, including the traditional data sources, as discussed previously. We aim to highlight the advantages and disadvantages of using each data source as a useful guide for future research on happiness.

With the growth of technology, researchers are inclined to use more innovative approaches for the measurement of happiness. In fact, over the last years, researchers use novel methodologies and data sources, which offer new opportunities to study happiness and to circumvent the limitations carried from traditional methodologies and data sources. Fowler and Christakis study [303] is one of the first and most important to help the transition of happiness research from the traditional to the innovative era. The researchers computerised information from archived handwritten administrative tracking sheets from the Framingham Heart Study. They study happiness as a network phenomenon, by using data of 4739 people, from 1983 to 2003. Comparing to previous traditional work on happiness, which main focus is on socioeconomic, political, and genetic factors, this study is the first one to study happiness as a spreading phenomenon and its characteristics. In particular, they suggest that happiness is a network phenomenon, which clusters happy and unhappy people and spreads across various social relationships (e.g., relatives, friends) up to three degrees of separation (e.g., to one's friends' friends' friends). Additionally, individuals that are central in the network are more likely to be happy in the future.

There are more than the study mentioned above in the innovative era, predominantly with the use of innovative big data sources. Although measuring happiness with new data approaches appears to be adequate in predicting the emotional component of happiness, most studies seem to neglect the structural component of happiness [304]. Below new data sources are described, and relevant studies are provided. We would like to underline that in comparison to objective well-being studies, researchers of subjective well-being usually explore more than one dimension.

Table 10 The table contains a subset of the information returned by a Twitter API

Id	Hashtags	Mentions	Text	Profile info
240556	#dinner #ny	[10214]	#dinner bihday your majesty @user #ny	{…..}
4261063	#lyft	[964215]	@user thanks for #lyft credit	{…..}
72096	null	null	factsguide society now	{…..}

Each tweet contains the information of the user profile and mentions or hashtags used in the text

3.2.1 Social media

Nowadays, people are highly involved in social media, and they are motivated to share their emotions and thoughts online, leaving a large and continuously updated user-generated content. Studying happiness from users' posts may eliminate the social desirability effect that traditional self-reports bring, due to participants' inaccurate and dishonest evaluation of happiness [305]. Thus, researchers and policy-makers are attracted by these intellectual opportunities to explore happiness, with wider use of Twitter data accessed through Twitter's public API. Twitter has the least barriers in data extraction, while the other social media have strict policies, and the acquisition of data has turned to be difficult. Social media data may also encounter some concerns. They may reflect social desirability biases since individuals manage their online profiles [122]. Also, Twitter users may not be as representative of the general population [123] as anonymized self-reports conducted through a chosen representative sample. Table 10 illustrates an example of the structure of Twitter records.

There are several studies on social media (mostly on Twitter) showing the variations on happiness as influenced by the universal needs (a2), and in particular, the interaction with other people. For example, Quercia et al. [306] use Twitter data in order to monitor the gross community happiness in the city of London. In particular, they suggest that Twitter friends, on average, have similar sentiment. They also show that the relationship between sentiment and well-being can hold at individual and community level. Bollen et al. [307] use the OpinionFinder (OF) subjectivity lexicon [308] in order to analyze the sentiment of an online social network of 39,110 Twitter users. They show the first direct observation of a significant Happiness Paradox, meaning that on average most of the individuals are less happy than their friends are. Similarly, by using the OF, Bollen et al. [309] analyze the emotional content of a set of Twitter users over 6 months, to examine whether happiness is assortative in online social networks. They find significant levels of happiness assortativity across Twitter, since users might be propense to connect to users with similar happiness values (homophilic attachment) or converge on their friends' happiness level (contagion). This result suggests that real social networks may work similarly. With the use of Facebook, Kramer et al. [310] test whether emotions are contagious between users without the awareness of the influenced individuals. Indeed, by reducing the amount of emotional content in the Facebook News Feed on an experiment conducted on Facebook users, they demonstrate that emotional contagion can also happen without direct interaction between the users and even without non-verbal cues.

Social media is also used for the exploration of happiness as influenced by the social environment dimension (a3). For example, Lim et al. [311] collect a set of geotagged tweets, of users in Melbourne, Australia, between the period of November 2016 to January 2017. They use sentiment analysis to demonstrate that people show more positive emotions and less negative emotions in green spaces or close to them. This could potentially be taken into consideration by policy-makers aiming to improve the societal well-being by urban greening interventions. Besides, Mitchell et al. [312] use Twitter data to study happiness and the 2010 United States Census Bureau's MAF/TIGER database to define the urban areas. They use the Language Assessment by Mechanical Turk (labMT) sentiment analysis tool to study the similarities in word use in urban areas in the United States, to map areas according to the happiness level and score individual states and cities for average word happiness. Golder and Macy [313] identify individual-level diurnal and seasonal mood rhythms in cultures across the globe, using data from Twitter between February 2008 and January 2010. They find that people like the weekend as people are much happier on Saturdays and Sundays. They also find that even individuals' good mood deteriorates as the day progresses, which is consistent with the effects of sleep and circadian rhythm. They also show that seasonal change in baseline positive affect varies with change in day length. Landsdall et al. [314] turn their attention to the issue of the public mood or sentiment—the mood of the nation. They use tweets sampled from the 54 largest cities in the UK from July 2009 to January 2012, and they associate each of the basic emotions (fear, joy, anger, sadness) with a list of words. They find out that each of the four key emotions changes over time in a manner that is partly predictable (or at least interpretable). Joy rises in Christmas, fear in Halloween, and especially negative mood started in October 2010, where massive cuts were announced in the UK. Cresci et al. [315] use Instagram data to explore, among others, the differences that the cultural and social environment bring on people's smiles. They perform face recognition in a case study of over 2 million selfies shared from January to February 2015. In particular, they use a Face++ algorithm function to measure the smiling degree of the individuals in their selfies. Results reveal that El Salvador, Brazil, and Panama have the highest smiling average.

Other researchers use social media to study the variations of happiness as influenced by more than one dimension. For example, Bollen et al. [316] conduct sentiment analysis on Twitter data from 2008. They find that events in the social and cultural (a3), political (a5), and economic sphere (a4) have a significant effect on happiness. Dodds et al. [317] construct the Hedonometer to measure temporal patterns of societal happiness, as influenced by basic needs (a2), as well as by various social (a3), economic (a4) and political (a5) determinants. For indicating happiness using Hedonometer, they create a data set of users' tweets over 3 years (from September 2008 to September 2011 approximately). The results show that in general, at an annual level, the average happiness appears to increase till April 2009 and then to decrease gradually. On a weekly basis, the average happiness peaks during the weekend and on an hourly basis, the happiest hour of the day is between 5 to 6 a.m. (US local time). Another example is Iacus et al. [318], who analyze tweets from Italy, written in the Italian language. In particular, they use the iSA (integrated Sentiment Analysis) method [319,320] to capture a set of determinants that influence happiness, such as self-esteem (a1) and family relationships (a2), and aggregate them into an index labeled SWBI (Social Well Being Index). Results suggest that the environmental and health conditions (a3) anticipate several determinants of happiness as measured by SWBI. This study is one of the few to study both the emotional and structural components of happiness. Curini et al. [321] use tweets posted in 2012 in Italy to build a happiness index, labeled iHappy. They demonstrate that variables such as the overall quality of institutions (a5) seem to have a minor effect on the average level of happiness of the Italian provinces. In contrast, meteorological variables, such as rain and snow (a3), as well as events related to specific days, such as the payday (a4), have a stronger impact on happiness. Furthermore, Durahim et al. [322] use Twitter data to create the Gross National Happiness (GNH) for the country of Turkey. The GNH created measures people's happiness as varied due to specific events, such as Saint Valentine's Day (a2), Starting day of Gezi Park Protests (#occupygezi), and Day of Ergenekon lawsuit verdict (a5). Last, Coviello et al. [323] compare what people post on Facebook to data they have on the weather (a3), specifically the rainfall amount. They find that people tend to post less happy messages on Facebook if it rains. This emotion seems to pass along their network (a2). For example, if a friend on Facebook is in a rainy area and this affects the emotional content of her posts on Facebook, then more likely, her friends might post a sadder message, even though where they are the weather is better.

3.2.2 Google trends

Another new data source is Google Trends, which provides data on the frequency of specific search terms over time. Algan et al. [324] present Google Trends as a new data source for exploring happiness and its relevant dimensions. They consider it a promising data source for its timeliness, since it provides computational social scientists with immediate data, as well as offers the possibility to observe people's behavior, as compared to analyzing textual opinions. On the other hand, working with Google Trends challenges researchers since the value of the series obtained directly from Google Trends is difficult to interpret, and this value on a given day cannot be compared between terms since they are normalized to the maximum value by term. In this study [324], researchers cover 300 weeks from January 6, 2008, to January 4, 2014. Results reveal that happiness is associated with job security, financial security (b4), family life (b2), and leisure determinants (b3). An example of Google Trends data set is not provided since data are represented as time series of the frequency.

3.2.3 Crowdsourced data

Crowdsourcing, as discussed in Sect. 2.2, involves obtaining work, information, or opinions from a large group of people who submit their data via the Internet, smartphone apps, etc. In particular, smartphones are lately appealing to happiness researchers since they give access to previously inaccessible data related to daily social behavior [325,326]. Innovative smartphone sensor technology, such as accelerometers, GPS, and Bluetooth, are used in combination with self-reports, such as mood tracking self-reports, in the form of EMA. However, such methodologies bring the limitations of the traditional data sources (see the first rows of Table 9), since happiness fluctuations are collected through self-reports. Moreover, when hiring individuals to participate in crowdsourcing platforms, the crowd is not anymore for free, and the study might result in high costs. It is, therefore, hard to keep a trade-off between initial objectives with results of quality and cost [327]. Additionally, some studies are conducted with a small number of data and might need to be replicated. Table 11 shows an example of crowdsourced data.

For example, Lathia et al. [328] collect data of over 10,000 individuals, by combining smartphone-based self-reports (in the form of EMA) and the accelerator in the smartphones, to investigate the relationship between happiness and physical activity (c3). Results show that there is indeed a relationship between happiness and physical activities, including the non-exercise ones, such as standing and walking. Asai et al. [329] study 100,000 happy moments from HappyDB over 3 months, to find which are the short and long term determinants of happiness. In particular, HappyDB is a database

Table 11 The table contains a subset of the information returned by HappyDB, a crowdsourced database capturing happy moments

Id	Reflection period	Text	Num. sentences
28775	24 h	Donated blood. Painful	2
32612	24 h	Morning yoga class	1
42663	24 h	Children with butterflies	1

created through Amazon Mechanical Turk, for capturing people's happy moments by asking every 24 h and once over 3 months, people's happiness status, and analyzing with NLP people's responses. Results show that exercise, nature, and leisure (c3) are short-term determinants, whereas social relationships with loved ones (c2) and achievements (c3) are long-term determinants. Bogomolov et al. [330] exploit a data set of 117 individuals, who are equipped with a sensing software between 2010 and 2011. This software collects smartphone activity data of call logs, SMS and proximity data (acquired by scanning nearby phones and other Bluetooth devices every five minutes). It also collects personality traits (the "Big Five" [331]) and daily happiness data by self-report questionnaires. Results demonstrate that by using mobile phone data reflecting social interactions (c2), information concerning weather conditions (c3), and personality traits (c1), individuals' daily happiness can be predicted.

3.2.4 News data

Similarly to objective well-being, news data are a new promising data source for the further exploration of subjective well-being. Its advantages and its disadvantages, as well as a data set example, are discussed and presented in Sect. 2.2. Carlquist et al. [332] study happiness with the use of news data. In particular, they study the concept of well-being in Norwegian society by examining word use patterns in four electronically archived Norwegian newspapers media from 1992 to 2014. They demonstrate that about half of the words referring to affective approaches, cognitive or life satisfaction approaches, eudaimonic and humanistic approaches, and character strengths show systematic and statistically significant patterns of change. The most notable rise concerns the eudaimonic words (related to mastery, motivation, and self-development), which show increasing trends in all newspapers. The authors state that certain happiness terms appearing more frequently could be interpreted as an increased and liberating focus on individual opportunity (d1) [333] or could demonstrate neoliberal ideology (d5) [334].

4 Discussion

In this study, we provide researchers with the theoretical background on both the objective and the subjective well-being or happiness, as well as their relevant dimensions necessary for the conduction of a meaningful study. In addition, we present a review of the data sources used for the exploration of well-being, and we discuss existing related studies. More specifically, we present the structure and the opportunities that each data source offers and the problems that researchers might encounter when working with these data.

The paper is primarily targeted at researchers interested in "Data Science for Social Good" (DS4SG) or similarly "Artificial Intelligence for Social Good" (AI4SG). Harnessed correctly, artificial intelligence can inform and empower the social good decision-making [335,336]. DS4SG or AI4SG is a vague concept, and there is not an adequate definition yet. However, Shi et al. [39] propose several societal application domains to shed light on this concept, such as healthcare and well-being. In this study, we specifically aim to contribute to the exploration of well-being through data science. Researchers from various disciplines, from social science to computer science, could use this paper to understand data science for well-being better and make a positive and tangible social impact.

We would like to underline that this is not a complete review of studies conducted on well-being with the use of innovative data sources. We aim to provide some examples of the most important evidence on these data sources and well-being dimensions so that this study works as a reference point for future research. We do not fully cover existing research on a given link that is present in Figs. 1 and 2, but to the best of our knowledge, a missing link entails that there is no existing study connecting the two nodes. For example, there is no adequate literature on news data for the exploration of the safety dimension (E5) of objective well-being. Therefore, since nowadays, safety is an important dimension, due to constant conflicts around the world (e.g., political instability, terrorist attacks), it shows great potential for future research.

Moreover, new data sources seem to be particularly promising for a more in-depth exploration of subjective well-being. Taking into consideration the subjective nature of happiness, it has been traditionally measured through self-reports. Although they have been proved to be valid, they are very costly, and depending on the study might neglect to capture either the emotional or the structural component of well-being. Therefore, new data sources could be used, and innovative methodologies, such as text analysis, could be applied for a complete, according to its definition, measurement of subjective well-being. Still, most studies using new data sources tap into the emotional component of subjective well-being and neglect the structural component. Consequently, we suggest further exploration of the novel data sources for the measurement of subjective well-being, capturing both components.

Undoubtedly, the research opportunities opened up by the innovative data sources discussed in this paper are plenty. However, with the use of these data sources, researchers are called to deal with new challenges comparing to traditional research. Since, usually, the data used are personal, if not sensitive, and are analyzed to shape policy and to make decisions [337,338], ethical concerns may arise, such as privacy and respect to human rights. In the European Union, additional attention to the topic has been brought after the implementation of the General Data Protection Regulation (GDPR). Researchers need to take into consideration the ethical challenges and not overlook them but address them successfully. Only by facing ethical problems, researchers can maximize the contributing value of data science studies for society.

Acknowledgements This work was supported by the European Commission through the Horizon2020 European project "SoBigData Research Infrastructure—Big Data and Social Mining Ecosystem" (Grant Agreement 654024). We would like to thank Daniele Fadda for support on data visualization.

Author contributions VV: conceptualization, writing, tables and figures, LG: conceptualization and writing, IM: writing, tables and figures, SC: writing, RS: writing, MT: writing, LP: conceptualization, writing and managing.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interests.

References

1. Reinhart, C.M., Reinhart, V.R.: After the fall. Technical report. National Bureau of Economic Research (2010)
2. Fleurbaey, M.: Beyond gdp: the quest for a measure of social welfare. J. Econ. Lit. **47**(4), 1029–75 (2009)
3. Stiglitz, J.E., Sen, A., Fitoussi, J.P.: Report by the Commission on the Measurement of Economic Performance and Social Progress. The Commission Paris (2009)
4. Dodge, R., Daly, A.P., Huyton, J., Sanders, L.D.: The challenge of defining wellbeing. Int. J. Wellbeing **2**(3), 11 (2012)
5. Alkire, S.: Dimensions of human development. World Dev. **30**(2), 181–205 (2002)
6. Organisation for Economic Co-operation and Development How's life? Measuring Well-Being. OECD, Paris (2011)
7. UNDP Sustainable Development Goals. https://sustainabledevelopment.un.org/sdgs. Accessed Oct 2019 (2015)
8. Rapporto, BES Il benessere equo e sostenibile in Italia. ISTAT (2015)
9. Organisation for Economic Co-operation and Development (OECD) OECD Guidelines on Measuring Subjective Well-Being. OECD Publishing (2013)
10. Veenhoven, R.: Conditions of Happiness, Reidel. Springer, Dordrecht (1984)
11. Frey, B.S., Stutzer, A.: What can economists learn from happiness research? J. Econ. Lit. **40**(2), 402–435 (2002)
12. Stiglitz, J.E., Sen, A., Fitoussi, J.P.: Measurement of economic performance and social progress. Online document. http://www.bitly/JTwmG Accessed 26 June 2012 (2009)
13. Bartels, M., Boomsma, D.I.: Born to be happy? The etiology of subjective well-being. Behav. Genet. **39**(6), 605 (2009)
14. Bartels, M., Saviouk, V., De Moor, M.H., Willemsen, G., van Beijsterveldt, T.C., Hottenga, J.J., De Geus, E.J., Boomsma, D.I.: Heritability and genome-wide linkage scan of subjective happiness. Twin Res. Hum. Genet. **13**(2), 135–142 (2010)
15. Nes, R.B., Røysamb, E.: The heritability of subjective well-being: review and meta-analysis. In: The Genetics of Psychological Well-Being: The Role of Heritability and Genetics in Positive Psychology, pp. 75–96 (2015)
16. Nes, R.B., Czajkowski, N., Tambs, K.: Family matters: happiness in nuclear families and twins. Behav. Genet. **40**(5), 577–590 (2010)
17. Nes, R., Røysamb, E., Tambs, K., Harris, J., Reichborn-Kjennerud, T.: Subjective well-being: genetic and environmental contributions to stability and change. Psychol. Med. **36**(7), 1033–1042 (2006)
18. Røysamb, E., Harris, J.R., Magnus, P., Vittersø, J., Tambs, K.: Subjective well-being. Sex-specific effects of genetic and environmental factors. Personal. Individ. Differ. **32**(2), 211–223 (2002)
19. Røysamb, E., Tambs, K., Reichborn-Kjennerud, T., Neale, M.C., Harris, J.R.: Happiness and health: environmental and genetic contributions to the relationship between subjective well-being, perceived health, and somatic illness. J. Pers. Soc. Psychol. **85**(6), 1136 (2003)
20. Schnittker, J.: Happiness and success: genes, families, and the psychological effects of socioeconomic position and social support. Am. J. Sociol. **114**(S1), S233–S259 (2008)
21. Pleeging, E., Burger, M., van Exel, J.: The relations between hope and subjective well-being: a literature overview and empirical analysis. Appl. Res. Qual. Life **1**, 1–23 (2020)
22. Kenrick, D.T., Griskevicius, V., Neuberg, S.L., Schaller, M.: Renovating the pyramid of needs: contemporary extensions built upon ancient foundations. Perspect. Psychol. Sci. **5**(3), 292–314 (2010)
23. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. **55**(1), 68 (2000)
24. Tay, L., Diener, E.: Needs and subjective well-being around the world. J. Pers. Soc. Psychol. **101**(2), 354 (2011)

25. Clark, A.E., Oswald, A.J.: Satisfaction and comparison income. J. Public Econ. **61**(3), 359–381 (1996)
26. Shields, M.A., Price, S.W., Wooden, M.: Life satisfaction and the economic and social characteristics of neighbourhoods. J. Popul. Econ. **22**(2), 421–443 (2009)
27. Powdthavee, N.: How much does money really matter? Estimating the causal effects of income on happiness. Empir. Econ. **39**(1), 77–92 (2010)
28. Nikolaev, B.: Living with mom and dad and loving it... or are you? J. Econ. Psychol. **51**, 199–209 (2015)
29. Dolan, P., Peasgood, T., White, M.: Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. J. Econ. Psychol. **29**(1), 94–122 (2008)
30. Easterlin, R.A.: Does economic growth improve the human lot? Some empirical evidence. In: Nations and Households in Economic Growth, pp 89–125. Elsevier (1974)
31. Veenhoven, R.: Is happiness relative? Soc. Indic. Res. **24**(1), 1–34 (1991)
32. Diener, E., Tay, L., Oishi, S.: Rising income and the subjective well-being of nations. J. Pers. Soc. Psychol. **104**(2), 267 (2013)
33. Veenhoven, R., Vergunst, F.: The Easterlin illusion: economic growth does go with greater happiness. Int. J. Happiness Dev. **1**(4), 311–343 (2014)
34. Sacks, D.W., Stevenson, B., Wolfers, J.: The new stylized facts about income and subjective well-being. Emotion **12**(6), 1181 (2012)
35. Radcliff, B., Shufeldt, G.: Direct democracy and subjective well-being: the initiative and life satisfaction in the American states. Soc. Indic. Res. **128**(3), 1405–1423 (2016)
36. Veenhoven, R.: Social conditions for human happiness: a review of research. Int. J. Psychol. **50**(5), 379–391 (2015)
37. Deaton, A.: The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. The World Bank (1997)
38. European Project.: SoBigData. http://sobigdata.eu/index. Accessed Oct 2019 (2015)
39. Shi, Z.R., Wang, C., Fang, F.: Artificial Intelligence for Social Good: A Survey. arXiv preprint arXiv:2001.01818 (2020)
40. Solomon, D.J.: Conducting web-based surveys. Pract. Assess. Res. Eval. **7**(19), 12 (2001)
41. Daas, P.J., Puts, M.J., Buelens, B., Van den Hurk, P.A.: Big data and official statistics. In: Proceedings of the NTTS, pp. 5–7. New Techniques and Technologies for Statistics (2013)
42. Struijs, P., Daas, P.: Quality approaches to big data in official statistics. In: European Conference on Quality in Official Statistics (2014)
43. Jahani, E., Sundsøy, P., Bjelland, J., Bengtsson, L., de Montjoye, Y.A., et al.: Improving official statistics in emerging markets using machine learning and mobile phone data. EPJ Data Sci. **6**(1), 3 (2017)
44. Blumenstock, J.E.: Fighting poverty with data. Science **353**(6301), 753–754 (2016)
45. United Nations.: A world that counts: mobilizing the data revolution for sustainable development. Technical report (2014)
46. Sustainable Development Solutions Network: Indicators and a Monitoring Framework for the Sustainable Development Goals. Launching a Data Revolution for the SDGs, United Nations, New York (2015)
47. WHO, World Health Organization: Geneva Macroeconomics and health: investing in health for economic development-report of the commission on macroeconomics and health. Commission on Macroeconomics and Health (2001)
48. European Commission: The Lisbon strategy for growth and jobs (2000)
49. OECD.: OECD Better Life Index: Health. http://www.oecdbetterlifeindex.org/topics/health/. Accessed Oct 2019 (2011)
50. OECD.: OECD Better Life Index: Jobs. http://www.oecdbetterlifeindex.org/topics/jobs/. Accessed Oct 2019 (2011a)
51. OECD.: OECD Better Life Index: Income. http://www.oecdbetterlifeindex.org/topics/income/. Accessed Oct 2019 (2011b)
52. OECD.: OECD Better Life Index: Environment. http://www.oecdbetterlifeindex.org/topics/environment/. Accessed Oct 2019 (2011c)
53. OECD.: OECD Better Life Index: Safety. http://www.oecdbetterlifeindex.org/topics/safety/. Accessed Oct 2019 (2011d)
54. Amerio, P., Roccato, M.: Psychological reactions to crime in Italy: 2002–2004. J. Commun. Psychol. **35**(1), 91–102 (2007)
55. OECD.: OECD Better Life Index: Civic Engagement. http://www.oecdbetterlifeindex.org/topics/civic-engagement/. Accessed Oct 2019 (2011)
56. Blondel, V.D., Decuyper, A., Krings, G.: A survey of results on mobile phone datasets analysis. EPJ Data Sci. **4**(1), 10 (2015)
57. Eagle, N., Pentland, A.S.: Eigenbehaviors: identifying structure in routine. Behav. Ecol. Sociobiol. **63**(7), 1057–1066 (2009)
58. Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barabási, A.L.: Returners and explorers dichotomy in human mobility. Nat. Commun. **6**, 8166 (2015)
59. Pappalardo, L., Rinzivillo, S., Simini, F.: Human mobility modelling: exploration and preferential return meet the gravity model. Proc. Comput. Sci. **83**, 934–939 (2016). https://doi.org/10.1016/j.procs.2016.04.188
60. Pellungrini, R., Pappalardo, L., Pratesi, F., Monreale, A.: A data mining approach to assess privacy risk in human mobility data. ACM Trans. Intell. Syst. Technol. **9**(3), 31:1–31:27 (2017). https://doi.org/10.1145/3106774
61. Pappalardo, L., Simini, F.: Data-driven generation of spatio-temporal routines in human mobility. Data Min. Knowl. Disc. **32**(3), 787–829 (2018)
62. Giannotti, F., Pappalardo, L., Pedreschi, D., Wang, D.: A Complexity Science Perspective on Human Mobility, pp. 297–314. Cambridge University Press, Cambridge (2013). https://doi.org/10.1017/CBO9781139128926.016
63. Ranjan, G., Zang, H., Zhang, Z.L., Bolot, J.: Are call detail records biased for sampling human mobility? ACM SIGMOBILE Mob. Comput. Commun. Rev. **16**(3), 33–44 (2012)
64. Iovan, C., Olteanu-Raimond, A.M., Couronné, T., Smoreda, Z,: Moving and calling: mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies. In: Geographic Information Science at the Heart of Europe, pp. 247–265. Springer (2013)
65. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature **453**(7196), 779 (2008)
66. Barabasi, A.L.: The origin of bursts and heavy tails in human dynamics. Nature **435**(7039), 207 (2005)
67. Oliver, N., Matic, A., Frias-Martinez, E.: Mobile network data for public health: opportunities and challenges. Front. Public Health **3**, 189 (2015)
68. Finger, F., Genolet, T., Mari, L., de Magny, G.C., Manga, N.M., Rinaldo, A., Bertuzzo, E.: Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. Proc. Nat. Acad. Sci. **113**(23), 6421–6426 (2016)
69. Kafsi, M., Kazemi, E., Maystre, L., Yartseva, L., Grossglauser, M., Thiran, P.: Mitigating epidemics through mobile micro-measures. arXiv preprint arXiv:1307.2084 (2013)
70. Lima, A., De Domenico, M., Pejovic, V., Musolesi, M.: Disease containment strategies based on mobility and information dissemination. Sci. Rep. **5**, 10650 (2015)
71. Madan, A., Cebrian, M., Lazer, D., Pentland, A.: Social sensing for epidemiological behavior change. In: Proceedings of the 12th

ACM International Conference on Ubiquitous Computing, pp. 291–300. ACM (2010)

72. Pappalardo, L., Pedreschi, D., Smoreda, Z., Giannotti, F.: Using big data to study the link between human mobility and socio-economic development. In: 2015 IEEE International Conference on Big Data (Big Data), pp. 871–78 (2015) https://doi.org/10.1109/BigData.2015.7363835
73. Toole, J.L., Lin, Y.R., Muehlegger, E., Shoag, D., González, M.C., Lazer, D.: Tracking employment shocks using mobile phone data. J. R. Soc. Interface **12**(107), 20150185 (2015)
74. Sundsøy, P., Bjelland, J., Reme, B.A., Jahani, E., Wetter, E., Bengtsson, L.: Towards real-time prediction of unemployment and profession. In: International Conference on Social Informatics, pp. 14–23. Springer (2017)
75. Eagle, N., Macy, M., Claxton, R.: Network diversity and economic development. Science **328**(5981), 1029–1031 (2010)
76. Steele, J.E., Sundsøy, P.R., Pezzulo, C., Alegana, V.A., Bird, T.J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.A., Iqbal, A.M., et al.: Mapping poverty using mobile phone and satellite data. J. R. Soc. Interface **14**(127), 20160690 (2017)
77. Mao, H., Shuai, X., Ahn, Y.Y., Bollen, J.: Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to côte d'ivoire. EPJ Data Sci. **4**(1), 15 (2015)
78. Gutierrez, T., Krings, G., Blondel, V.D.: Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets. arXiv preprint arXiv:1309.4496 (2013)
79. Blumenstock, J.: Calling for better measurement: estimating an individual's wealth and well-being. ACM KDD (Data Mining for Social Good) (2014)
80. Blumenstock, J., Cadamuro, G., On, R.: Predicting poverty and wealth from mobile phone metadata. Science **350**(6264), 1073–1076 (2015)
81. Frias-Martinez, V., Virseda, J.: On the relationship between socio-economic factors and cell phone usage. In: Proceedings of the Fifth International Conference on Information and Communication Technologies and Development, pp. 76–84. ACM (2012)
82. Soto, V., Frias-Martinez, V., Virseda, J., Frias-Martinez, E.: Prediction of socioeconomic levels using cell phone records. In: International Conference on User Modeling, Adaptation, and Personalization, pp. 377–388. Springer (2011)
83. Frias-Martinez, V., Soguero-Ruiz, C., Frias-Martinez, E., Josephidou, M.: Forecasting socioeconomic trends with cell phone records. In: Proceedings of the 3rd ACM Symposium on Computing for Development, p. 15. ACM (2013)
84. Hernandez, M., Hong, L., Frias-Martinez, V., Frias-Martinez, E.: Estimating poverty using cell phone data: evidence from Guatemala. The World Bank (2017)
85. Pappalardo, L., Vanhoof, M., Gabrielli, L., Smoreda, Z., Pedreschi, D., Giannotti, F.: An analytical framework to now-cast well-being using mobile phone data. Int. J. Data Sci. Anal. **2**(1), 75–92 (2016). https://doi.org/10.1007/s41060-016-0013-2
86. Lotero, L., Cardillo, A., Hurtado, R., Gómez-Gardeñes, J.: Several multiplexes in the same city: the role of socioeconomic differences in urban mobility. In: Interconnected Networks, pp. 149–164. Springer (2016)
87. Amini, A., Kung, K., Kang, C., Sobolevsky, S., Ratti, C.: The impact of social segregation on human mobility in developing and industrialized regions. EPJ Data Sci. **3**(1), 6 (2014)
88. Smith-Clarke, C., Mashhadi, A., Capra, L.: Poverty on the cheap: estimating poverty maps using aggregated mobile communication networks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 511–520. , ACM (2014)
89. Picornell, M., Ruiz, T., Borge, R., García-Albertos, P., de la Paz, D., Lumbreras, J.: Population dynamics based on mobile phone data to improve air pollution exposure assessments. J. Expos. Sci. Environ. Epidemiol. **29**(2), 278 (2019)
90. Lu, X., Wrathall, D.J., Sundsøy, P.R., Nadiruzzaman, M., Wetter, E., Iqbal, A., Qureshi, T., Tatem, A.J., Canright, G.S., Engø-Monsen, K., et al.: Detecting climate adaptation with mobile network data in bangladesh: anomalies in communication, mobility and consumption patterns during cyclone mahasen. Clim. Change **138**(3–4), 505–519 (2016)
91. Lu, X., Bengtsson, L., Holme, P.: Predictability of population displacement after the 2010 haiti earthquake. Proc. Nat. Acad. Sci. **109**(29), 11576–11581 (2012)
92. Bengtsson, L., Lu, X., Thorson, A., Garfield, R., Von Schreeb, J.: Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in haiti. PLoS Med. **8**(8), e1001083 (2011)
93. Wilson, R., Zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., Guthrie, S., Chamberlain, H., Brooks, C., Hughes, C., et al.: Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 Nepal earthquake. PLoS Curr. **8**, 1 (2016)
94. Nyarku, M., Mazaheri, M., Jayaratne, R., Dunbabin, M., Rahman, M.M., Uhde, E., Morawska, L.: Mobile phones as monitors of personal exposure to air pollution: Is this the future? PLoS ONE **13**(2), e0193150 (2018)
95. Liu, H.Y., Skjetne, E., Kobernus, M.: Mobile phone tracking: in support of modelling traffic-related air pollution contribution to individual exposure and its implications for public health impact assessment. Environ. Health **12**(1), 93 (2013)
96. Decuyper, A., Rutherford, A., Wadhwa, A., Bauer, J.M., Krings, G., Gutierrez, T., Blondel, V.D., Luengo-Oroz, M.A.: Estimating food consumption and poverty indices with mobile phone data. arXiv preprint arXiv:1412.2595 (2014)
97. Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., Pentland, A.: Once upon a crime: towards crime prediction from demographics and mobile data. In: Proceedings of the 16th International Conference on Multimodal Interaction, pp. 27–434. ACM (2014)
98. Ferrara, E., De Meo, P., Catanese, S., Fiumara, G.: Detecting criminal organizations in mobile phone networks. Expert Syst. Appl. **41**(13), 5733–5750 (2014)
99. Elgethun, K., Fenske, R.A., Yost, M.G., Palcisko, G.J.: Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument. Environ. Health Perspect. **111**(1), 115–122 (2003)
100. Dias, D., Tchepel, O.: Modelling of human exposure to air pollution in the urban environment: a GPS-based approach. Environ. Sci. Pollut. Res. **21**(5), 3558–3571 (2014)
101. Beekhuizen, J., Kromhout, H., Huss, A., Vermeulen, R.: Performance of gps-devices for environmental exposure assessment. J. Eposure Sci. Environ. Epidemiol. **23**(5), 498 (2013)
102. Pappalardo, L., Simini, F., Barlacchi, G., Pellungrini, R.: Scikit-mobility: a python library for the analysis, generation and risk assessment of mobility data. arXiv:1907.07062 (2019)
103. Jankowska, M.M., Schipperijn, J., Kerr, J.: A framework for using GPS data in physical activity and sedentary behavior studies. Exerc. Sport Sci. Rev. **43**(1), 48 (2015)
104. Kelly, P., Krenn, P., Titze, S., Stopher, P., Foster, C.: Quantifying the difference between self-reported and global positioning systems-measured journey durations: a systematic review. Transp. Rev. **33**(4), 443–459 (2013)
105. Meurs, H., Haaijer, R.: Spatial structure and mobility. Transp. Res. Part D Transp. Environ. **6**(6), 429–446 (2001)
106. Oliver, M., Badland, H., Mavoa, S., Duncan, M.J., Duncan, S.: Combining GPS, GIS, and accelerometry: methodological issues

in the assessment of location and intensity of travel behaviors. J. Phys. Activity Health **7**(1), 102–108 (2010)

107. Adams, S.A., Matthews, C.E., Ebbeling, C.B., Moore, C.G., Cunningham, J.E., Fulton, J., Hebert, J.R.: The effect of social desirability and social approval on self-reports of physical activity. Am. J. Epidemiol. **161**(4), 389–398 (2005)
108. Pappalardo, L., Rinzivillo, S., Qu, Z., Pedreschi, D., Giannotti, F.: Understanding the patterns of car travel. Eur. Phys. J. Spec. Top. **215**(1), 61–73 (2013). https://doi.org/10.1140/epjst/e2013-01715-5
109. Chaix, B., Kestens, Y., Duncan, D.T., Brondeel, R., Méline, J., El Aarbaoui, T., Pannier, B., Merlo, J.: A GPS-based methodology to analyze environment-health associations at the trip level: case-crossover analyses of built environments and walking. Am. J. Epidemiol. **184**(8), 579–589 (2016)
110. Kerr, J., Duncan, S., Schipperjin, J.: Using global positioning systems in health research: a practical approach to data collection and processing. Am. J. Prev. Med. **41**(5), 532–540 (2011)
111. Saelens, B.E., Vernez Moudon, A., Kang, B., Hurvitz, P.M., Zhou, C.: Relation between higher physical activity and public transit use. Am. J. Public Health **104**(5), 854–859 (2014)
112. Rundle, A.G., Sheehan, D.M., Quinn, J.W., Bartley, K., Eisenhower, D., Bader, M.M., Lovasi, G.S., Neckerman, K.M.: Using GPS data to study neighborhood walkability and physical activity. Am. J. Prev. Med. **50**(3), e65–e72 (2016)
113. Sadler, R.C., Gilliland, J.A.: Comparing children's GPS tracks with geospatial proxies for exposure to junk food. Spat. Spat. Temp. Epidemiol. **14**, 55–61 (2015)
114. Canzian, L., Musolesi, M.: Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 1293–1304. ACM (2015)
115. Marchetti, S., Giusti, C., Pratesi, M., Salvati, N., Giannotti, F., Pedreschi, D., Rinzivillo, S., Pappalardo, L., Gabrielli, L.: Small area model-based estimators using big data sources. J. Off. Stat. **31**(2), 263–281 (2015)
116. Smith, C., Quercia, D., Capra, L.: Finger on the pulse: identifying deprivation using transit flow analysis. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work, pp. 683–692. ACM (2013)
117. Lathia, N., Quercia, D., Crowcroft, J.: The hidden image of the city: sensing community well-being from urban mobility. In: International Conference on Pervasive Computing, pp. 91–98. Springer (2012)
118. Robinson, A.I., Carnes, F., Oreskovic, N.M.: Spatial analysis of crime incidence and adolescent physical activity. Prev. Med. **85**, 74–77 (2016)
119. Ariel, B., Partridge, H.: Predictable policing: measuring the crime control benefits of hotspots policing at bus stops. J. Quant. Criminol. **33**(4), 809–833 (2017)
120. Spinsanti, L., Berlingerio, M., Pappalardo, L.: Mobility and Geo-Social Networks, pp. 315–333. Cambridge University Press, Cambridge (2013). https://doi.org/10.1017/CBO9781139128926.017
121. Olteanu, A., Castillo, C., Diaz, F., Kiciman, E.: Social data: biases, methodological pitfalls, and ethical boundaries. Front. Big Data **2**, 13 (2019)
122. Rost, M., Barkhuus, L., Cramer, H., Brown, B.: Representation and communication: challenges in interpreting large social media datasets. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work, pp. 357–362. ACM (2013)
123. Eichstaedt, J.C., Schwartz, H.A., Kern, M.L., Park, G., Labarthe, D.R., Merchant, R.M., Jha, S., Agrawal, M., Dziurzynski, L.A., Sap, M., et al.: Psychological language on twitter predicts county-level heart disease mortality. Psychol. Sci. **26**(2), 159–169 (2015)
124. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. ICWSM **13**, 1–10 (2013)
125. Signorini, A., Segre, A.M., Polgreen, P.M.: The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PLoS ONE **6**(5), e19467 (2011)
126. Paul, M.J., Dredze, M., Broniatowski, D.: Twitter improves influenza forecasting. PLoS Curr. **6**, 12 (2014)
127. Lampos, V., Cristianini, N.: Tracking the flu pandemic by monitoring the social web. In: 2010 2nd International Workshop on Cognitive Information Processing, pp. 411–416. IEEE (2010)
128. Lampos, V., Cristianini, N.: Nowcasting events from the social web with statistical learning. ACM Trans. Intell. Syst. Technol. **3**(4), 72 (2012)
129. Chen, X., Yang, X.: Does food environment influence food choices? A geographical analysis through "tweets". Appl. Geogr. **51**, 82–89 (2014)
130. Llorente, A., Garcia-Herranz, M., Cebrian, M., Moro, E.: Social media fingerprints of unemployment. PLoS ONE **10**(5), e0128692 (2015)
131. Antenucci, D., Cafarella, M., Levenstein, M., Ré, C., Shapiro, M.D.: Using social media to measure labor market flows. Technical report. National Bureau of Economic Research (2014)
132. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. J. Comput. Sci. **2**(1), 1–8 (2011)
133. Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., Goldstein, G. Identifying and following expert investors in stock microblogs. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp 1310–1319. Association for Computational Linguistics (2011)
134. De Choudhury, M., Sundaram, H., John, A., Seligmann, D.D.: Can blog communication dynamics be correlated with stock market activity? In: Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia, pp. 55–60. ACM (2008)
135. Cresci, S., Lillo, F., Regoli, D., Tardelli, S., Tesconi, M.: $FAKE: Evidence of spam and bot activity in stock microblogs on Twitter. In: Proceedings of the 12th International Conference on Web and Social Media (ICWSM'18), pp. 580–583. AAAI (2018)
136. Cresci, S., Lillo, F., Regoli, D., Tardelli, S., Tesconi, M.: Cashtag piggybacking: uncovering spam and bot activity in stock microblogs on twitter. ACM Trans. Web (TWEB) **13**(2), 11 (2019)
137. Avvenuti, M., Cresci, S., Marchetti, A., Meletti, C., Tesconi, M.: Predictability or early warning: using social media in modern emergency response. IEEE Internet Comput. **20**(6), 4–6 (2016)
138. Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., Cebrian, M.: Rapid assessment of disaster damage using social media activity. Sci. Adv. **2**(3), e1500779 (2016)
139. Avvenuti, M., Cresci, S., La Polla, M.N., Meletti, C., Tesconi, M.: Nowcasting of earthquake consequences using big social data. IEEE Internet Comput. **6**, 37–45 (2017)
140. Mendoza, M., Poblete, B., Valderrama, I.: Nowcasting earthquake damages with twitter. EPJ Data Sci. **8**(1), 3 (2019)
141. Avvenuti, M., Cresci, S., Del Vigna, F., Tesconi, M.: Impromptu crisis mapping to prioritize emergency response. Computer **49**(5), 28–37 (2016)
142. Avvenuti, M., Cresci, S., Del Vigna, F., Fagni, T., Tesconi, M.: CrisMap: a big data crisis mapping system based on damage detection and geoparsing. Inf. Syst. Front. **1**, 1–19 (2018)
143. Preis, T., Moat, H.S., Bishop, S.R., Treleaven, P., Stanley, H.E.: Quantifying the digital traces of hurricane sandy on flickr. Sci. Rep. **3**, 3141 (2013)
144. Chen, X., Cho, Y, Jang, S.Y.: Crime prediction using twitter sentiment and weather. In: 2015 Systems and Information Engineering Design Symposium, pp. 63–68. IEEE (2015)

145. Al Boni, M., Gerber, M.S.: Predicting crime with routine activity patterns inferred from social media. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 001233–001238. IEEE (2016)
146. Kadar, C., Brüngger, R.R., Pletikosa, I.: Measuring ambient population from location-based social networks to describe urban crime. In: International Conference on Social Informatics, pp. 521–535. Springer (2017)
147. Chen, F., Neill, D.B.: Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1166–1175. ACM (2014)
148. Nobles, M., Neill, D.B., Flaxman, S.: Predicting and Preventing Emerging Outbreaks of Crime (2014)
149. Neill, D.B., Gorr, W.L.: Detecting and preventing emerging epidemics of crime. Adv. Dis. Surveill. **4**(13), 18 (2007)
150. Colleoni, E., Rozza, A., Arvidsson, A.: Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. J. Commun. **64**(2), 317–332 (2014)
151. Goh, T.T., Xin, Z., Jin, D.: Habit formation in social media consumption: a case of political engagement. Behav. Inf. Technol. **38**(3), 273–288 (2019)
152. Ferrara, E.: Manipulation and abuse on social media. ACM SIGWEB Newsl. **2015**(Spring), 4 (2015)
153. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: The paradigm-shift of social spambots: evidence, theories, and tools for the arms race. In: Proceedings of the 26th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, pp 963–972 (2017)
154. Goldstein, B.A., Navar, A.M., Pencina, M.J., Ioannidis, J.: Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J. Am. Med. Inform. Assoc. **24**(1), 198–208 (2017)
155. Wilson, P.W., D'Agostino, R.B., Levy, D., Belanger, A.M., Silbershatz, H., Kannel, W.B.: Prediction of coronary heart disease using risk factor categories. Circulation **97**(18), 1837–1847 (1998)
156. Sultana, J., Leal, I., de Wilde, M., de Ridder, M., van der Lei, J., Sturkenboom, M., et al.: Identifying data elements to measure frailty in a dutch nationwide electronic medical record database for use in postmarketing safety evaluation: an exploratory study. Drug Saf. **12**, 1–7 (2019)
157. Ghaderighahfarokhi, S., Sadeghifar, J.: A model to predict low birth weight infants and affecting factors using data mining techniques. J. Basic Res. Med. Sci. **5**(3), 1–8 (2018)
158. Metzger, M.H., Tvardik, N., Gicquel, Q., Bouvry, C., Poulet, E., Potinet-Pagliaroli, V.: Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a french pilot study. Int. J. Methods Psychiatric Res. **26**(2), e1522 (2017)
159. Mhaskar, H.N., Pereverzyev, S.V., van der Walt, M.D.: A deep learning approach to diabetic blood glucose prediction. Front. Appl. Math. Stat. **3**, 14 (2017)
160. Santillana, M., Nsoesie, E.O., Mekaru, S.R., Scales, D., Brownstein, J.S.: Using clinicians' search query data to monitor influenza epidemics. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. **59**(10), 1446 (2014)
161. Althoff, T., Hicks, J.L., King, A.C., Delp, S.L., Leskovec, J., et al.: Large-scale physical activity data reveal worldwide activity inequality. Nature **547**(7663), 336 (2017)
162. Hayeri, A.: Predicting future glucose fluctuations using machine learning and wearable sensor data. Diabetes (2018). https://doi.org/10.2337/db18-738-P
163. Leetaru, K.: The GDELT Project. https://www.gdeltproject.org/. Accessed Oct 2019 (2013)
164. Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen, B., Belyaeva, J.: Sentiment analysis in the news. arXiv preprint arXiv:1309.6202 (2013)
165. Dehghan, A., Montgomery, L., Arciniegas-Mendez, M., Ferman-Guerra, M.: Predicting News Bias (2016)
166. Grein, T.W., Kamara, K., Rodier, G., Plant, A.J., Bovier, P., Ryan, M.J., Ohyama, T., Heymann, D.L.: Rumors of disease in the global village: outbreak verification. Emerg. Infect. Dis. **6**(2), 97 (2000)
167. Heymann, D.L., Rodier, G.R., et al.: Hot spots in a wired world: Who surveillance of emerging and re-emerging infectious diseases. Lancet. Infect. Dis **1**(5), 345–353 (2001)
168. Brownstein, J.S., Freifeld, C.C., Reis, B.Y., Mandl, K.D.: Surveillance sans frontieres: Internet-based emerging infectious disease intelligence and the healthmap project. PLoS Med. **5**(7), e151 (2008)
169. Wilson, K., Brownstein, J.S.: Early detection of disease outbreaks using the internet. CMAJ **180**(8), 829–831 (2009)
170. Chunara, R., Andrews, J.R., Brownstein, J.S.: Social and news media enable estimation of epidemiological patterns early in the 2010 haitian cholera outbreak. Am. J. Trop. Med. Hyg. **86**(1), 39–45 (2012)
171. Alanyali, M., Moat, H.S., Preis, T.: Quantifying the relationship between financial news and the stock market. Sci. Rep. **3**, 3578 (2013)
172. Lillo, F., Miccichè, S., Tumminello, M., Piilo, J., Mantegna, R.N.: How news affects the trading behaviour of different categories of investors in a financial market. Quant. Finance **15**(2), 213–229 (2015)
173. Kleinschmit, D., Sjöstedt, V.: Between science and politics: Swedish newspaper reporting on forests in a changing climate. Environ. Sci. Policy **35**, 117–127 (2014)
174. Boykoff, M.T.: Lost in translation? united states television news coverage of anthropogenic climate change, 1995–2004. Clim. Change **86**(1–2), 1–11 (2008)
175. Van Aelst, P., De Swert, K.: Politics in the News: Do Campaigns Matter? A Comparison of Political News During Election Periods and Routine Periods in Flanders (Belgium). Walter de Gruyter GmbH & Co, KG, Belgium (2009)
176. Eurostat Practical Guide for Processing Supermarket Scanner Data (2017)
177. Griffith, R., O'Connell, M.: The use of scanner data for research into nutrition. Fiscal Stud. **30**(3–4), 339–365 (2009)
178. Baron, S., Lock, A.: The challenges of scanner data. J. Oper. Res. Soc. **46**(1), 50–61 (1995)
179. Eurostat Practical Guide for Processing Supermarket Scanner Data. https://circabc.europa.eu/sd/a/8e1333df-ca16-40fc-bc6a-1ce1be37247c/Practical-Guide-Supermarket. Accessed Oct 2019 (2017)
180. Diewert, W.E.: Harmonized indexes of consumer prices: their conceptual foundations (2002)
181. Magruder, S.: Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of human disease. Johns Hopkins Univ. APL Tech. Dig. **24**(4), 349–353 (2003)
182. Bonnet, C., Dubois, P., Réquillart, V.: The dynamics of satured fat consumption in france. Technical. report. Toulouse mimeo (2008)
183. Griffith, R., Leibtag, E., Leicester, A., Nevo, A.: Consumer shopping behavior: how much do consumers save? J. Econ. Perspect. **23**(2), 99–120 (2009)
184. Janssen, A., Parslow, E.: Pregnancy and alcohol purchases: evidence from scanner data. Avail. SSRN **3446559**, 12 (2019)
185. Rider, J., Berck, P., Villas-Boas, S.B.: Eating Healthy in Lean Times: The Relationship Between Unemployment and Grocery Purchasing Patterns (2012)

186. Van der Grient, H.A., de Haan, J.: The use of supermarket scanner data in the dutch cpi. In: Joint ECE/ILO Workshop on Scanner Data, vol. 10 (2010)
187. Silver, M., Heravi, S.: Scanner data and the measurement of inflation. Econ. J. **111**(472), 383–404 (2001)
188. Pennacchioli, D., Coscia, M., Rinzivillo, S., Giannotti, F., Pedreschi, D.: The retail market as a complex system. EPJ Data Sci. **3**(1), 33 (2014)
189. Sobolevsky, S., Massaro, E., Bojic, I., Arias, J.M., Ratti, C.: Predicting regional economic indices using big data of individual bank card transactions. In: 2017 IEEE International Conference on Big Data (Big Data), pp. 1313–1318. IEEE (2017)
190. Panzone, L.A., Wossink, A., Southerton, D.: The design of an environmental index of sustainable food consumption: a pilot study using supermarket data. Ecol. Econ. **94**, 44–55 (2013)
191. Gadema, Z., Oglethorpe, D.: The use and usefulness of carbon labelling food: a policy perspective from a survey of uk supermarket shoppers. Food Policy **36**(6), 815–822 (2011)
192. Brancoli, P., Rousta, K., Bolton, K.: Life cycle assessment of supermarket food waste. Resour. Conserv. Recycl. **118**, 39–46 (2017)
193. Scholz, K., Eriksson, M., Strid, I.: Carbon footprint of supermarket food waste. Resour. Conserv. Recycl. **94**, 56–65 (2015)
194. Goel, S., Hofman, J.M., Lahaie, S., Pennock, D.M., Watts, D.J.: Predicting consumer behavior with web search. Proc. Nat. Acad. Sci. **107**(41), 17486–17490 (2010)
195. Cooper, C.P., Mallon, K.P., Leadbetter, S., Pollack, L.A., Peipins, L.A.: Cancer internet search activity on a major search engine, united states 2001–2003. J. Med. Internet Res. **7**(3), e36 (2005)
196. Polgreen, P.M., Chen, Y., Pennock, D.M., Nelson, F.D., Weinstein, R.A.: Using internet searches for influenza surveillance. Clin. Infect. Dis. **47**(11), 1443–1448 (2008)
197. Hulth, A., Rydevik, G., Linde, A.: Web queries as a source for syndromic surveillance. PLoS ONE **4**(2), e4378 (2009)
198. Yuan, Q., Nsoesie, E.O., Lv, B., Peng, G., Chunara, R., Brownstein, J.S.: Monitoring influenza epidemics in china with search query from baidu. PLoS ONE **8**(5), e64323 (2013)
199. Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L.: Detecting influenza epidemics using search engine query data. Nature **457**(7232), 1012 (2009)
200. Google: Google Flu Trends. http://www.google.org/flutrends. Accessed Oct 2019 (2008)
201. Nsoesie, E., Mararthe, M., Brownstein, J.: Forecasting peaks of seasonal influenza epidemics. PLoS Curr. **5**, 8 (2013)
202. Yang, W., Lipsitch, M., Shaman, J.: Inference of seasonal and pandemic influenza transmission dynamics. Proc. Nat. Acad. Sci. **112**(9), 2723–2728 (2015)
203. Wilson, N., Mason, K., Tobias, M., Peacey, M., Huang, Q., Baker, M.: Interpreting “google flu trends” data for pandemic h1n1 influenza: the new zealand experience. Eurosurveillance **14**(44), 19386 (2009)
204. Chan, E.H., Sahai, V., Conrad, C., Brownstein, J.S.: Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Neglect. Trop. Dis. **5**(5), e1206 (2011)
205. Althouse, B.M., Ng, Y.Y., Cummings, D.A.: Prediction of dengue incidence using search query surveillance. PLoS Neglect. Trop. Dis. **5**(8), e1258 (2011)
206. Dukic, V.M., David, M.Z., Lauderdale, D.S.: Internet queries and methicillin-resistant staphylococcus aureus surveillance. Emerg. Infect. Dis. **17**(6), 1068 (2011)
207. Ocampo, A.J., Chunara, R., Brownstein, J.S.: Using search queries for malaria surveillance, Thailand. Malaria J. **12**(1), 390 (2013)
208. Yang, A.C., Tsai, S.J., Huang, N.E., Peng, C.K.: Association of internet search trends with suicide death in taipei city, taiwan, 2004–2009. J. Affect. Disord. **132**(1–2), 179–184 (2011)
209. McCarthy, M.J.: Internet monitoring of suicide risk in the population. J. Affect. Disord. **122**(3), 277–279 (2010)
210. Kristoufek, L., Moat, H.S., Preis, T.: Estimating suicide occurrence statistics using google trends. EPJ Data Sci. **5**(1), 32 (2016)
211. Adler, N., Cattuto, C., Kalimeri, K., Paolotti, D., Tizzoni, M., Verhulst, S., Yom-Tov, E., Young, A.: How search engine data enhance the understanding of determinants of suicide in india and inform prevention: observational study. J. Med. Internet Res. **21**(1), e10179 (2019). https://doi.org/10.2196/10179
212. Ettredge, M., Gerdes, J., Karuga, G.: Using web-based search data to predict macroeconomic statistics. Commun. ACM **48**(11), 87–92 (2005)
213. Askitas, N., Zimmermann, K.: Google econometrics and unemployment forecasting. Appl. Econ. Quart. **55**(2), 107–120 (2009)
214. Francesco/FD D, Marcucci J “google it!” forecasting the us unemployment rate with a google job search index. Mpra paper. University Library of Munich, Germany. https://EconPapers.repec.org/RePEc:pra:mprapa:18248 (2009)
215. Suhoy, T., et al.: Query indices and a 2008 downturn: Israeli data. Technical report. Bank of Israel (2009)
216. Baker, S., Fradkin, A., et al.: What drives job search? evidence from google search data. Discussion Papers, pp. 10–20 (2011)
217. McLaren, N., Shanbhogue, R.: Using internet search data as economic indicators. Bank Engl. Quart. Bull. **51**(2), 134–140 (2011)
218. Choi, H., Varian, H.: Predicting initial claims for unemployment benefits. Google Inc, pp. 1–5 (2009)
219. Choi, H., Varian, H.: Predicting the present with google trends. Econ. Rec. **88**, 2–9 (2012)
220. Koop, G., Onorante, L.: Macroeconomic nowcasting using google probabilities. In: First International Conference on Advanced Research Methods and Analytics, CARMA2016. https://doi.org/10.4995/CARMA2016.2016.4213 (2016)
221. Guzman, G.: Internet search behavior as an economic forecasting tool: the case of inflation expectations. J. Econ. Soc. Meas. **36**(3), 119–167 (2011)
222. Preis, T., Reith, D., Stanley, H.E.: Complex dynamics of our economic life on different scales: insights from search engine query data. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **368**(1933), 5707–5719 (2010). https://doi.org/10.1098/rsta.2010.0284
223. Preis, T., Moat, H.S., Stanley, H.E.: Quantifying trading behavior in financial markets using google trends. Sci. Rep. (2013). https://doi.org/10.1038/srep01684
224. Curme, C., Preis, T., Stanley, H.E., Moat, H.S.: Quantifying the semantics of search behavior before stock market moves. Proc. Natl. Acad. Sci. **111**(32), 11600–11605 (2014). https://doi.org/10.1073/pnas.1324054111
225. Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., Weber, I.: Web search queries can predict stock market volumes. PLoS ONE **7**(7), e40014 (2012)
226. Moat, H.S., Curme, C., Avakian, A., Kenett, D.Y., Stanley, H.E., Preis, T.: Quantifying wikipedia usage patterns before stock market moves. Sci. Rep. **3**, 1801 (2013)
227. Qi, H., Manrique, P., Johnson, D., Restrepo, E., Johnson, N.F.: Open source data reveals connection between online and on-street protest activity. EPJ Data Sci. **5**(1), 18 (2016a)
228. Qi, H., Manrique, P., Johnson, D., Restrepo, E., Johnson, N.F.: Association between volume and momentum of online searches and real-world collective unrest. Results Phys. **6**, 414–419 (2016b)
229. Chykina, V., Crabtree, C.: Using google trends to measure issue salience for hard-to-survey populations. Socius **4**, 2378023118760414 (2018)
230. Reilly, S., Richey, S., Taylor, J.B.: Using google search data for state politics research: an empirical validity test using roll-off data. State Polit. Policy Quart. **12**(2), 146–159 (2012)

231. Kleemann, F., Voß, G.G., Rieder, K.: Un (der) paid innovators: the commercial utilization of consumer work through crowdsourcing. Sci. Technol. Innov. Stud. **4**(1), 5–26 (2008)
232. Behrend, T.S., Sharek, D.J., Meade, A.W., Wiebe, E.N.: The viability of crowdsourcing for survey research. Behav. Res. Methods **43**(3), 800 (2011)
233. Paolotti, D., Carnahan, A., Colizza, V., Eames, K., Edmunds, J., Gomes, G., Koppeschaar, C., Rehn, M., Smallenburg, R., Turbelin, C., et al.: Web-based participatory surveillance of infectious diseases: the influenzanet participatory surveillance experience. Clin. Microbiol. Infect. **20**(1), 17–21 (2014)
234. Dalton, C., Durrheim, D., Fejsa, J., Francis, L., Carlson, S., d'Espaignet, E.T., Tuyl, F., et al.: Flutracking: a weekly australian community online survey of influenza-like illness in 2006, 2007 and 2008. Commun. Dis. Intell. Quart. Rep. **33**(3), 316 (2009)
235. Smolinski, M.S., Crawley, A.W., Baltrusaitis, K., Chunara, R., Olsen, J.M., Wójcik, O., Santillana, M., Nguyen, A., Brownstein, J.S.: Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. Am. J. Public Health **105**(10), 2124–2130 (2015)
236. Hashemian, M., Knowles, D., Calver, J., Qian, W., Bullock, MC., Bell, S., Mandryk, R.L., Osgood, N., Stanley, K.G.: iepi: an end to end solution for collecting, conditioning and utilizing epidemiologically relevant data. In: Proceedings of the 2nd ACM International Workshop on Pervasive Wireless Healthcare. pp. 3–8. ACM (2012)
237. Madan, A., Cebrian, M., Moturu, S., Farrahi, K., et al.: Sensing the "health state" of a community. IEEE Pervasive Comput. **11**(4), 36–45 (2011)
238. Martinucci, I., Natilli, M., Lorenzoni, V., Pappalardo, L., Monreale, A., Turchetti, G., Pedreschi, D., Marchi, S., Barale, R., de Bortoli, N.: Gastroesophageal reflux symptoms among italian university students: epidemiology and dietary correlates using automatically recorded transactions. BMC Gastroenterol. **18**(1), 116 (2018)
239. Green, T.C., Huang, R., Wen, Q., Zhou, D.: Crowdsourced employer reviews and stock returns. J. Financ. Econ. **2**, 18 (2019)
240. Dabirian, A., Kietzmann, J., Diba, H.: A great place to work!? understanding crowdsourced employer branding. Bus. Horiz. **60**(2), 197–205 (2017)
241. Könsgen, R., Schaarschmidt, M., Ivens, S., Munzel, A.: Finding meaning in contradiction on employee review sites-effects of discrepant online reviews on job application intentions. J. Interact. Mark. **43**, 165–177 (2018)
242. Tingzon, I., Orden, A., Sy, S., Sekara, V., Weber, I., Fatehkia, M., Herranz, M.G., Kim, D.: Mapping Poverty in the Philippines Using Machine Learning, Satellite Imagery, and Crowd-sourced Geospatial Information (missing year)
243. OpenStreetMap Community Openstreetmap. https://www.openstreetmap.org/#map=5/42.088/12.564. Accessed Oct 2019 (2004)
244. Piaggesi, S., Gauvin, L., Tizzoni, M., Cattuto, C., Adler, N., Verhulst, S., Young, A., Price, R., Ferres, L., Panisson, A.: Predicting city poverty using satellite imagery. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 90–96 (2019)
245. Abelson, B., Varshney, K.R., Sun, J.: Targeting direct cash transfers to the extremely poor. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1563–1572. ACM (2014)
246. Hersman, E., Okolloh, O., Rotich, J., Kobia, D.: Ushahidi. https://www.ushahidi.com. Accessed Oct 2019 (2008)
247. Meier, P.: Digital Humanitarians: How Big Data is Changing the Face of Humanitarian Response. Routledge, London (2015)
248. European Commission Citizens' Observatories. https://www.ushahidi.com. Accessed Oct 2019 (2016)
249. Grainger, A.: Citizen observatories and the new earth observation science. Remote Sens. **9**(2), 153 (2017)
250. Schneider, P., Castell, N., Vogt, M., Lahoz W., Bartonova A.: Making sense of crowdsourced observations: data fusion techniques for real-time mapping of urban air quality. In: EGU General Assembly Conference Abstracts, p. 17 (2015)
251. Meier, F., Fenner, D., Grassmann, T., Jänicke, B., Otto, M., Scherer, D.: Challenges and benefits from crowd sourced atmospheric data for urban climate research using Berlin, Germany, as testbed. In: ICUC9–9th International Conference on Urban Climate jointly with 12th Symposium on the Urban Environment (2015)
252. Chapman, L., Bell, C., Bell, S.: Can the crowdsourcing data paradigm take atmospheric science to a new level? a case study of the urban heat island of london quantified using netatmo weather stations. Int. J. Climatol. **37**(9), 3597–3605 (2017)
253. Lea, S.G., D'Silva, E., Asok, A.: Women's strategies addressing sexual harassment and assault on public buses: an analysis of crowdsourced data. Crime Prev. Commun. Saf. **19**(3–4), 227–239 (2017)
254. Gosselt, J.F., Van Hoof, J.J., Gent, B.S., Fox, J.P.: Violent frames: analyzing internet movie database reviewers' text descriptions of media violence and gender differences from 39 years of us action, thriller, crime, and adventure movies. Int. J. Commun. **9**, 547–567 (2015)
255. Ozkan, T., Worrall, J.L., Zettler, H.: Validating media-driven and crowdsourced police shooting data: a research note. J. Crime Justice **41**(3), 334–345 (2018)
256. Avvenuti, M., Bellomo, S., Cresci, S., La Polla, M.N., Tesconi, M.: Hybrid crowdsensing: A novel paradigm to combine the strengths of opportunistic and participatory crowdsensing. In: Proceedings of the 26th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, pp. 1413–1421 (2017)
257. Dennis, J.: United by what divides us: 38 degrees and the eu referendum. In: EU Referendum Analysis 2016: Media, Voters and the Campaign. Bournemouth University, p. 100 (2016)
258. Yasseri, T., Bright, J.: Wikipedia traffic data and electoral prediction: towards theoretically informed models. EPJ Data Sci. **5**(1), 22 (2016)
259. Gellers, J.C.: Crowdsourcing global governance: sustainable development goals, civil society, and the pursuit of democratic legitimacy. Int. Environ. Agreements Polit. Law Econ. **16**(3), 415–432 (2016)
260. Burger, R.: Aristotle's Dialogue with Socrates: On the "Nicomachean Ethics". University of Chicago Press, Chicago (2009)
261. Diener, E.: Subjective well-being. Psychol. Bull. **95**(3), 542 (1984)
262. Veenhoven, R.: How do we assess how happy we are? tenets, implications and tenability of three theories. Happiness Econ. Polit. **25**, 45–69 (2009)
263. Alesina, A., Di Tella, R., MacCulloch, R.: Inequality and happiness: are europeans and americans different? J. Public Econ. **88**(9–10), 2009–2042 (2004)
264. Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. **54**(6), 1063 (1988)
265. Watson, D., Clark, L.A.: The Panas-x: Manual for the Positive and Negative Affect Schedule-Expanded Form. Psychology Publications, New York (1999)
266. Diener, E., Oishi, S., Tay, L.: Advances in subjective well-being research. Nat. Hum. Behav. **2**, 1 (2018)
267. Hudson, N.W., Anusic, I., Lucas, R.E., Donnellan, M.B.: Comparing the reliability and validity of global self-report measures of subjective well-being with experiential day reconstruction measures. Assessment **2**, 26 (2017)

268. Anusic, I., Schimmack, U.: Stability and change of personality traits, self-esteem, and well-being: introducing the meta-analytic stability and change model of retest correlations. J. Pers. Soc. Psychol. **110**(5), 766 (2016)
269. Tay, L., Chan, D., Diener, E.: The metrics of societal happiness. Soc. Indic. Res. **117**(2), 577–600 (2014)
270. Deaton, A.: Income, health, and well-being around the world: evidence from the gallup world poll. J. Econ. Perspect. **22**(2), 53–72 (2008)
271. Easterlin, R.A., Angelescu, L.: Happiness and growth the world over: time series evidence on the happiness-income paradox. Technical report. Institute of Labor Economics (IZA) (2009)
272. Kahneman, D., Deaton, A.: High income improves evaluation of life but not emotional well-being. Proc. Nat. Acad. Sci. **107**(38), 16489–16493 (2010)
273. Frijters, P., Beatton, T.: The mystery of the u-shaped relationship between happiness and age. J. Econ. Behav. Organ. **82**(2–3), 525–542 (2012)
274. Stevenson, B., Wolfers, J.: The paradox of declining female happiness. Am. Econ. J. Econ. Policy **1**(2), 190–225 (2009)
275. Deaton, A., Stone, A.A.: Understanding context effects for a measure of life evaluation: how responses matter. Oxf. Econ. Pap. **68**(4), 861–870 (2016)
276. Yap, S.C., Wortman, J., Anusic, I., Baker, S.G., Scherer, L.D., Donnellan, M.B., Lucas, R.E.: The effect of mood on judgments of subjective well-being: nine tests of the judgment model. J. Pers. Soc. Psychol. **113**(6), 939 (2017)
277. Lucas, R.E., Lawless, N.M.: Does life seem better on a sunny day? Examining the association between daily weather conditions and life satisfaction judgments. J. Pers. Soc. Psychol. **104**(5), 872 (2013)
278. Kahneman, D., Diener, E., Schwarz, N.: Well-Being: Foundations of Hedonic Psychology. Russell Sage Foundation, New York (1999)
279. Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., Stone, A.A.: A survey method for characterizing daily life experience: the day reconstruction method. Science **306**(5702), 1776–1780 (2004)
280. Courvoisier, D.S., Eid, M., Lischetzke, T.: Compliance to a cell phone-based ecological momentary assessment study: the effect of time and personality characteristics. Psychol. Assess. **24**(3), 713 (2012)
281. Shiffman, S., Stone, A.A., Hufford, M.R.: Ecological momentary assessment. Annu. Rev. Clin. Psychol. **4**, 1–32 (2008)
282. Eid, M.E., Diener, E.E.: Handbook of Multimethod Measurement in Psychology. American Psychological Association, New York (2006)
283. Diener, E., Seligman, M.E.: Beyond money: toward an economy of well-being. Psychol. Sci. Public Interest **5**(1), 1–31 (2004)
284. Costa, P.T., McCrae, R.R.: Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people. J. Pers. Soc. Psychol. **38**(4), 668 (1980)
285. Zweig, J.S.: Are women happier than men? Evidence from the Gallup World Poll. J. Happiness Stud. **16**(2), 515–541 (2015)
286. Deaton, A.S., Tortora, R.: People in Sub-Saharan Africa rate their health and health care among the lowest in the world. Health Aff. **34**(3), 519–527 (2015)
287. Veenhoven, R., Ehrhardt, J.: The cross-national pattern of happiness: test of predictions implied in three theories of happiness. Soc. Indic. Res. **34**(1), 33–68 (1995)
288. Cuñado, J., de Gracia, F.P.: Does education affect happiness? Evidence for spain. Soc. Indic. Res. **108**(1), 185–196 (2012)
289. Nikolaev, B.: Does higher education increase hedonic and eudaimonic happiness? J. Happiness Stud. **19**(2), 483–504 (2018)
290. Rehdanz, K., Maddison, D.: Climate and happiness. Ecol. Econ. **52**(1), 111–125 (2005)
291. Hudson, J.: Institutional trust and subjective well-being across the eu. Kyklos **59**(1), 43–62 (2006)
292. Hayo, B. Happiness in Eastern Europe. Marburg Economic Working Paper No 12 (2004)
293. Ferrer-i Carbonell, A., Gowdy, J.M.: Environmental degradation and happiness. Ecol. Econ. **60**(3), 509–516 (2007)
294. Gardner, J., Oswald, A.J.: Money and mental wellbeing: a longitudinal study of medium-sized lottery wins. J. Health Econ. **26**(1), 49–60 (2007)
295. Tay, L., Zyphur, M., Batz, C.: Income and Subjective Well-Being: Review, Synthesis, and Future Research. Handbook of Well-Being. DEF Publishers, Salt Lake City (2017)
296. Wijngaards, I., Hendriks, M., Burger, M.J.: Steering towards happiness: an experience sampling study on the determinants of happiness of truck drivers. Transp. Res. Part A Policy Pract. **128**, 131–148 (2019)
297. van der Zwan, P., Hessels, J., Burger, M.: Happy free willies? Investigating the relationship between freelancing and subjective well-being. Small Bus. Econ. **8**, 1–17 (2019)
298. Blanchflower, D.G., Bell, D.N., Montagnoli, A., Moro, M.: The happiness trade-off between unemployment and inflation. J. Money Credit Bank. **46**(S2), 117–141 (2014)
299. Knabe, A., Schöb, R., Weimann, J.: Partnership, gender, and the well-being cost of unemployment. Soc. Indic. Res. **129**(3), 1255–1275 (2016)
300. Brulé, G., Veenhoven, R.: Why are Latin Europeans less happy? Polyphonic Anthropology-Theoretical and Empirical Cross-Cultural Fieldwork. The Impact of Hierarchy. InTech (2012)
301. Bartolini, S., Mikucka, M., Sarracino, F.: Money, trust and happiness in transition countries: evidence from time series. Soc. Indic. Res. **130**(1), 87–106 (2017)
302. Ott, J.C.: Good governance and happiness in nations: technical quality precedes democracy and quality beats size. J. Happiness Stud. **11**(3), 353–368 (2010)
303. Fowler, J.H., Christakis, N.A.: Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. BMJ **337**, a2338 (2008)
304. Luhmann, M.: Using big data to study subjective well-being. Curr. Opin. Behav. Sci. **18**, 28–33 (2017)
305. Nederhof, A.J.: Methods of coping with social desirability bias: a review. Eur. J. Soc. Psychol. **15**(3), 263–280 (1985)
306. Quercia, D., Ellis, J., Capra, L., Crowcroft, J.: Tracking gross community happiness from tweets. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, pp. 965–968. ACM (2012)
307. Bollen, J., Gonçalves, B., van de Leemput, I., Ruan, G.: The happiness paradox: your friends are happier than you. EPJ Data Sci. **6**(1), 4 (2017)
308. Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: OpinionFinder: a system for subjectivity analysis. In: Proceedings of hlt/emnlp on Interactive Demonstrations. Association for Computational Linguistics, pp. 34–35 (2005)
309. Bollen, J., Gonçalves, B., Ruan, G., Mao, H.: Happiness is assortative in online social networks. Artif. Life **17**(3), 237–251 (2011)
310. Kramer, A.D., Guillory, J.E., Hancock, J.T.: Experimental evidence of massive-scale emotional contagion through social networks. In: Proceedings of the National Academy of Sciences, p. 201320040 (2014)
311. Lim, K.H., Lee, K.E., Kendal, D., Rashidi, L., Naghizade, E., Winter, S., Vasardani, M.: The grass is greener on the other side: Understanding the effects of green spaces on twitter user sentiments. In: Companion of the The Web Conference 2018 on The Web Conference 2018, International World Wide Web Conferences Steering Committee, pp. 275–282 (2018)

312. Mitchell, L., Frank, M.R., Harris, K.D., Dodds, P.S., Danforth, C.M.: The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place. PLoS ONE **8**(5), e64417 (2013)
313. Golder, S.A., Macy, M.W.: Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science **333**(6051), 1878–1881 (2011)
314. Lansdall-Welfare, T., Lampos, V., Cristianini, N.: Nowcasting the mood of the nation. Significance **9**(4), 26–28 (2012)
315. Cresci, S., La Polla, M.N., Mazza, M., Tesconi, M., Del Vigna, F.: #selfie: mapping the phenomenon. Consiglio Nazioonale delle Ricerche IIT TR-08/2016 Technical Report (2016)
316. Bollen, J., Mao, H., Pepe, A.: Modeling public mood and emotion: twitter sentiment and socio-economic phenomena. ICWSM **11**, 450–453 (2011)
317. Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.A., Danforth, C.M.: Temporal patterns of happiness and information in a global social network: hedonometrics and twitter. PLoS ONE **6**(12), e26752 (2011)
318. Iacus, S.M., Porro, G., Salini, S., Siletti, E.: Social networks, happiness and health: from sentiment analysis to a multidimensional indicator of subjective well-being. arXiv preprint arXiv:1512.01569 (2015)
319. Ceron, A., Curini, L., Iacus, S.M.: Social Media e Sentiment Analysis: L'evoluzione dei fenomeni sociali attraverso la Rete, vol. 9. Springer, New York (2014)
320. Ceron, A., Curini, L., Iacus, S.M.: ISA: a fast, scalable and accurate algorithm for sentiment analysis of social media content. Inf. Sci. **367**, 105–124 (2016)
321. Curini, L., Iacus, S., Canova, L.: Measuring idiosyncratic happiness through the analysis of twitter: an application to the italian case. Soc. Indic. Res. **121**(2), 525–542 (2015)
322. Durahim, A.O., Coşkun, M.: # iamhappybecause: gross national happiness through twitter analysis and big data. Technol. Forecast. Soc. Change **99**, 92–105 (2015)
323. Coviello, L., Sohn, Y., Kramer, A.D., Marlow, C., Franceschetti, M., Christakis, N.A., Fowler, J.H.: Detecting emotional contagion in massive social networks. PLoS ONE **9**(3), e90315 (2014)
324. Algan, Y., Beasley, E., Guyot, F., Higa, K., Murtin, F., Senik, C., et al. Big Data Measures of Well-Being: Evidence from a Google Well-Being Index in the United States. OECD Statistics Working Papers 2016 (2016)
325. Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T.: A survey of mobile phone sensing. IEEE Commun. Mag. **48**(9), 140–150 (2010)
326. Staiano, J., Lepri, B., Aharony, N., Pianesi, F., Sebe, N., Pentland, A.: Friends don't lie: inferring personality traits from social network structure. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 321–330. ACM (2012)
327. Li, G., Zheng, Y., Fan, J., Wang, J., Cheng, R.: Crowdsourced data management: overview and challenges. In: Proceedings of the 2017 ACM International Conference on Management of Data, pp. 1711–1716. ACM (2017)
328. Lathia, N., Sandstrom, G.M., Mascolo, C., Rentfrow, P.J.: Happier people live more active lives: using smartphones to link happiness and physical activity. PLoS ONE **12**(1), e0160589 (2017)
329. Asai, A., Evensen, S., Golshan, B., Halevy, A., Li, V., Lopatenko, A., Stepanov, D., Suhara, Y., Tan, W.C., Xu, Y. Happydb: a corpus of 100,000 crowdsourced happy moments. arXiv preprint arXiv:1801.07746 (2018)
330. Bogomolov, A., Lepri, B., Pianesi, F.: Happiness recognition from mobile phone data. In: Social Computing (SocialCom), 2013 International Conference on Social Computing, pp. 790–795. IEEE (2013)
331. Goldberg, L.R.: An alternative "description of personality": the big-five factor structure. J. Pers. Soc. Psychol. **59**(6), 1216 (1990)
332. Carlquist, E., Nafstad, H.E., Blakar, R.M., Ulleberg, P., Delle Fave, A., Phelps, J.M.: Well-being vocabulary in media language: an analysis of changing word usage in Norwegian newspapers. J. Positive Psychol. **12**(2), 99–109 (2017)
333. Seligman, M.E.: Flourish: A New Understanding of Happiness and Well-Being and How to Achieve Them. Nicholas Brealey, Boston (2011)
334. Greco, M., Stenner, P.: Happiness and the art of life: diagnosing the psychopolitics of wellbeing. Health Cult. Soc. **5**(1), 1–19 (2013)
335. Coulton, C.J., Goerge, R., Putnam-Hornstein, E., de Haan, B.: Harnessing Big Data for Social Good: A Grand Challenge for Social Work, pp. 1–20. American Academy of Social Work and Social Welfare, Cleveland (2015)
336. Lepri, B., Staiano, J., Sangokoya, D., Letouzé, E., Oliver, N.: The tyranny of data? The bright and dark sides of data-driven decision-making for social good. In: Transparent Data Mining for Big and Small Data, pp. 3–24. Springer (2017)
337. Floridi, L., Taddeo, M.: What is data ethics? The Royal Society (2016)
338. Hand, D.J.: Aspects of data ethics in a changing world: where are we now? Big Data **6**(3), 176–190 (2018)