

Rieckmann, Johannes; Schanze, Jan-Lucas

## Article

# Perceptions of personal security in social media and search engines: A realistic reflection of actual crime rates?

DIW Economic Bulletin

## Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Rieckmann, Johannes; Schanze, Jan-Lucas (2015) : Perceptions of personal security in social media and search engines: A realistic reflection of actual crime rates?, DIW Economic Bulletin, ISSN 2192-7219, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin, Vol. 5, Iss. 12, pp. 178-186

This Version is available at:

<https://hdl.handle.net/10419/108853>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Perceptions of Personal Security in Social Media and Search Engines— a Realistic Reflection of Actual Crime Rates?

By Johannes Rieckmann and Jan-Lucas Schanze

The most common method of measuring subjective fear of crime in the general population has traditionally been through surveys. With the spread of digital technologies, however, data from social media and search engines could now help researchers learn more about people's subjective perceptions of certain types of crime. The present article will show that although the analysis of data from social media and search engines is not suitable as an indicator of actual crime levels in Germany, it can certainly be a cost-effective supplement to traditional methods of collecting data on perceived crime levels.

The number of Internet users in Germany—around 80 percent of the entire population now use the Internet at least occasionally<sup>1</sup>—is steadily increasing. This makes the Internet an ever more attractive source of data for crime researchers. Traditionally, the main source of data about the impact of crime on the public has been the police crime statistics (*Polizeiliche Kriminalstatistik, PKS*) of the Federal Criminal Police Office (*Bundeskriminalamt*), supported by occasional and until now largely regional surveys on the “dark figure” of crime (unreported crime) and fear of crime. Online data analysis, on the other hand, could provide deeper insight into the perception of crime; in the long term, it could even become an alternative and independent source of data for crime researchers. The present article<sup>2</sup> will evaluate whether data collected on the Internet are suitable for supplementing—or even functioning as a cost-effective substitute for—traditional studies on the fear of crime. It will examine the search engine behavior of Internet users in Germany as well as the statements they make on various social media platforms. It is expressly not aimed at forecasting crime trends for the purposes of preventing and fighting crime (predictive policing). Rather, the user-oriented media analysis presented here is intended as an initial descriptive characterization of German perceptions of personal security. The findings

<sup>1</sup> B. Van Eimeren and B. Frees, “Ergebnisse der ARD/ZDF-Onlinestudie 2014. 79 Prozent der Deutschen online – Zuwachs bei mobiler Internetnutzung und Bewegtbild,” *Media Perspektiven*, no. 7–8 (2014). A 12,000-person survey conducted by the WISIND project concluded that the figure was approximately 75 percent.

<sup>2</sup> The report was compiled as part of the research project An Economic Security Indicator for Germany (WISIND), which is backed by the German Ministry for Education and Research as part of the Social Dimensions of Security Research funding program. The idea behind the WISIND project and the generation of WISIND-specific data was jointly developed by Martin Kroh, Mathias Bug, Kristina Meier, Johannes Rieckmann, Eric van Um, and Nina Wald, together with the staff of the Brandenburg Institute for Society and Security (Brandenburgisches Institut für Gesellschaft und Sicherheit, BIGS). The authors would also like to thank Enrique Fernandez, Martina Kraus and Bartosz Walenda for their assistance throughout this process.

will then be compared with fear of crime as measured by the WISIND project.<sup>3</sup>

### Social Media and Search Engines as Data Sources

The present article uses two types of data. First it draws on data on the number of social media posts about certain types of crime, specifically posts made on Facebook, Twitter, discussion forums, and blogs, as well as comments made on YouTube videos. The purpose of these posts is to express perceptions and opinions, and to communicate; their primary function is therefore expressive.

The second source of data used in the article comes from Germany's leading search engine, Google. Search engines primarily have an exploratory function, as their purpose is to help users acquire information. Google has created the Google Trends platform for exporting data on the occurrence of search terms for given time periods and locations. Google Trends has been used in several scientific studies to observe data generated in the lead-up to various phenomena<sup>4</sup> and it is used here as a source of search engine data.

### Motivation of Internet Users

While search engines are primarily used to acquire information, social media is used mainly to exchange information and views. The use of both, search engines and social media, implies that the user is personally affected, either directly or indirectly. It can be assumed that a large part of the population uses the Internet to gather information on what they perceive to be threatening events or circumstances; that some of them communicate this information on social media platforms; that they are also interested in taking precautions (for example, using alarm systems or pepper spray); and that, in the process, data is generated that creates a picture of public threat perception which is independent of surveys. However, any hypotheses regarding the specific motivations of users are speculative because while tracking data on the Internet shows what users are doing, it does not show why they are doing it.<sup>5</sup>

### Collecting Data from Social Media Platforms

DIW Berlin subcontracted data collection from social media platforms to Beck et al. Services GmbH, a company specializing in capturing data from such sources. The data were collected over a period of four and a half months, from June 12 to October 31, 2014. The data used for the analysis comprise all posts and profile details that were shared with the public by users on the media platforms mentioned above; all content was therefore publicly accessible. German-language posts on the five platforms were automatically searched every day using a Web crawler and a list of search terms for ten different crime categories.<sup>6</sup> The selection of terms, the inclusion of alternative spellings and the use of so-called killer terms (terms that cause content to be excluded) ensured that the results were highly accurate and the data free of irrelevant content. For the crime category "bodily harm," for example, all newly entered text fragments with the keywords *Schlägerei* (brawl), *verprügelt* (beaten up), and *Körperverletzung* (assault or physical injury) were counted. Content with the keyword *Gewalt* (violence) that was connected to places outside of Germany (Ukraine or Iraq, for example) or to abstract concepts, like *höhere Gewalt* (force majeure), were identified as irrelevant and excluded.

Only publicly available content was examined for the presence of search terms. From the start, it was technically impossible to record any messages or content on profile pages that could not be viewed by the general public. In order to be able to map the data by region, any place names mentioned in the text fragments were recorded, and searches were made for mentions of place names by users. For example, the mention of Berlin in someone's profile ("lives in Berlin") can help determine the location of an Internet post on bodily harm as well as the occurrence of a place name in the post itself ("bodily harm in Berlin").<sup>7</sup> The results of an automated identification of the tone of posts (positive, negative, or neutral) proved to be of only limited use for empirical analysis and are therefore not addressed in this article. No further data were collected on the Internet users, such as age, gender, or other characteristics. User-generated data were anonymized, fully safeguarding user privacy;

<sup>3</sup> See also the article by M. Bug, M. Kraus, and B. Walenda in this issue of DIW Economic Bulletin.

<sup>4</sup> One prominent example is Google Flu Trends (GFT) which was created to predict spikes in flu activity in the US. When people who may be infected with the flu use a search engine to look for (or "google") flu symptoms, the frequency of these queries within a particular region or time is detected by Google Trends and used to infer imminent surges in infection rates, even before patients go to the doctor. For a critique of GFT, see D. Lazer et al., "The Parable of Google Flu: Traps in Big Data Analysis," *Science* 343 (2014): 1203-1205.

<sup>5</sup> M. Mahrt and M. Scharnow, "The Value of Big Data in Digital Media Research," *Journal of Broadcasting & Electronic Media* 57, no. 1 (2013): 20-33.

<sup>6</sup> Categories were created for the following types of crime: internet crime, theft and burglary, drug-related crime, bodily harm, organized crime, politically motivated crime, robbery, religious fundamentalism, sex crimes, and homicide.

<sup>7</sup> Place names in a post tend to be a more reliable indicator of where the incident under discussion occurred.

## Box

**Problems of Data Collection on the Internet**

The analysis of data from social media, often referred to as “big data,” has been the focus of increasing criticism, with charges that the approach taken by some research projects is purely data-driven and not grounded in theory.<sup>1</sup> The mere availability and size of new data sources is not a sufficient argument for the necessity of research. On the contrary, research must always identify problems and qualitative issues and, if possible, establish connections to “traditional” data collection.

The first and most obvious drawback to working with data generated on the Internet is that the data are not representative. Although the group of Internet users in Germany has now become very large numerically, it presents an image that is systematically distorted in favor of younger generations. In the “over-60s” generation, fewer than half of all Germans use the Internet. This deviation in the population is even greater when it comes to social media usage. Three-quarters of respondents under the age of 30 use social media, but only five percent of Internet users over 70 do so.<sup>2</sup>

In the case of Google Trends, it is not so much the pool of data itself as the way it is presented which is problematic. Google is a commercial enterprise and not a professional supplier of data for research or scientific purposes, and this is reflected in the lack of transparency in the internal processes used for data generation in Google Trends. The way the results are presented is

unfavorable for scientific analyses: they are shown not in terms of absolute numbers of searches but as relative search volumes, always in relation to a maximum value which is set at 100. According to Google, this normalization of the search volume is a result of expressing the search term as a fraction of all searches in a region, making it possible to compare results with other regions. If the total number of searches for a certain term is less than a threshold defined by Google but not publicly stated, the results are either not shown at all (search volume = 0), or they are shown only for longer time intervals (on a monthly instead of a weekly basis). This complex procedure is not made public by Google. Another major problem is that non-verifiable changes may be made to Google algorithms over time, making it more difficult to replicate data.<sup>3</sup> Changes in search volumes, particularly when they are observed over long periods of time,<sup>4</sup> are not always reliable evidence of an actual change in search behavior.

An additional distortion results when certain topics experience a temporary increase in media attention, which artificially drives up searches—a problem that “Google Flu Trends” also struggles with.<sup>5</sup> On the other hand, attempts to influence the search behavior of Internet users, for example because of the commercial interests of companies that want more clicks on their own pages and hope to sell products, are relatively unlikely in the field of internal security and crime—at least for the search terms used in this study.

**1** M. Welker and A. Klob, “Soziale Medien als Gegenstand und Instrument sozialwissenschaftlicher Forschung,” in *Soziale Medien – Gegenstand und Instrument der Forschung*, ed. C. König, M. Stahl, and E. Wiegand (Wiesbaden: Springer VS, 2014).

**2** Van Eimeren and Frees, *Ergebnisse der ARD/ZDF-Onlinestudie*: 380, 387.

**3** D. Lazer et al., “The Parable of Google Flu,” 1205.

**4** Google Trends provides data going back to 2004.

**5** D. Lazer et al., “The Parable of Google Flu,” 1204.

the sample does not contain any personal data or data that could be used to identify persons.<sup>8</sup>

### Collecting Search Engine Data (Google Trends)

The analysis of search engine usage was restricted to the market leader, Google. The Google search engine was used for 95 percent of all searches in Germany in

December 2014.<sup>9</sup> The use of search engines among the 55.6 million Internet users in Germany is widespread at 82 percent, according to the 2014 ARD/ZDF online study.<sup>10</sup> Analyzing data using Google Trends is therefore tantamount to collecting almost all search engine queries in Germany. The queries are aggregated by Google at different geographic levels and are not localized by the researchers themselves.<sup>11</sup>

**8** On data protection in the analysis of social media data, see R. Tscherwinka, “Soziale Medien – Gegenstand und Instrument der Forschung – Rechtliche Aspekte,” in *Soziale Medien – Gegenstand und Instrument der Forschung*, ed. C. König, M. Stahl, and E. Wiegand (Wiesbaden: Springer VS, 2014).

**9** Internet source: <http://de.statista.com/statistik/daten/studie/167841/umfrage/marktanteile-ausgewaehlter-suchmaschinen-in-deutschland/>

**10** Van Eimeren and Frees, “Ergebnisse der ARD/ZDF-Onlinestudie,” 387.

**11** There is no detailed information on the exact localization mechanisms used by Google. Presumably this is based on users’ IP addresses (i.e., the addresses of the users making queries).

Collecting data using Google Trends is a much simpler process than capturing data from social media. The user interface has a number of features that make it possible to compare results for different search terms, countries or time periods. Users can obtain regional information about Germany at the state level and filter results by choosing from among categories. The data sample used for this article was filtered by the category “Law and Government” to exclude potentially irrelevant results<sup>12</sup> stemming from the use of ambiguous search terms (see box).

## Findings

A comparison of all posts recorded on all social media platforms during the collection period shows that some issues occupied users’ attention significantly more than others (see Table 1). There were almost 300,000 posts on theft and burglary. These were followed closely by posts on religious fundamentalism; posts on homicide, bodily harm and drug-related crime trailed far behind. A comparison of the crime categories that users were interested in with PKS and the findings of the WISIND study on unreported crime supports the supposition that social networks cannot, by themselves, be used as a reliable means of inferring real regional incidence levels. The number of times one crime category was mentioned relative to other crime categories does not reflect the actual incidence level of this crime in Germany. It does, however provide an indication of general sentiment and perceived risk. For example, media coverage of the Hogesa demonstrations in Cologne at the end of October 2014, and of the rioting that accompanied them, led to a sharp rise across Germany in the occurrence of terms that were counted as keywords for the risk categories “religious fundamentalism” and “politically motivated crime.” (Hogesa is an abbreviation of the German phrase meaning “Hooligans against Salafis.”)

Of the 1.2 million posts collected for use in the present study, so far it has been possible to localize at least 18 percent in over 7,300 localities down to the municipal level and assign them to administrative districts in Germany. The results were adjusted for the numbers of inhabitants in the respective areas and converted to posts per 100,000 people.<sup>13</sup>

Table 1

### Post frequencies by offences in social media<sup>1</sup>

Offence	All posts	Among those: Localized	Proportion of localized offences
Theft and burglary	293,038	57,275	19.5
Religious fundamentalism	288,643	42,672	14.8
Bodily harm	180,133	37,011	20.5
Homicides	147,627	21,604	14.6
Drug related offences	109,260	16,155	14.8
Robbery	107,366	31,253	29.1
Internet crime	53,280	2,639	5.0
Sexual offences	43,217	3,770	8.7
Politically motivated crime	40,384	8,947	22.2
Organized crime	20,478	3,362	16.4
<b>Sum</b>	<b>1,283,426</b>	<b>224,688</b>	<b>17.5</b>

<sup>1</sup> June 12 through Oct. 31 2014.

Source: Bug, M.; Kroh, M.; Meier, K.; Rieckmann, J.; van Um, E.; Wald, N. (2015): WISIND-datasets: social media/Google Trends. Calculations by DIW Berlin.

© DIW Berlin 2015

The number one topics getting the most attention by social media users are theft and burglary as well as religious fundamentalism.

The degree to which content could be localized varied according to crime category. Users writing about robbery, political crimes, and burglary and theft mentioned place names more frequently in their posts than users writing about internet crime or sex crimes. Intuitively this can be explained by the fact that the physical location<sup>14</sup> is largely irrelevant for internet crime, and that specifying locations thus has little value in discussions about this type of crime. In the case of sex crimes, factors such as feelings of shame, frequent lack of witnesses, a sense of respect for the victim, and a desire to protect them presumably account for posts being less frequently associated with physical locations. These examples indicate a systemic imbalance in the social media data that poses a problem for any comparison of localized posts from social media platforms and actual regional crime statistics: some crimes can be assigned to a region more easily than others.

When analyzing social media, it is important to select multiple networks and sources, as selecting a single source further restricts the variance of the data.<sup>15</sup> Most of the posts in the data sample are from Facebook, the most widely used social network in Germa-

<sup>12</sup> Irrelevant results might include references to literature and film (for example, “Murder on the Orient Express”) when searching for murder, or results about illnesses such as Ebola when searching for virus (as in computer virus).

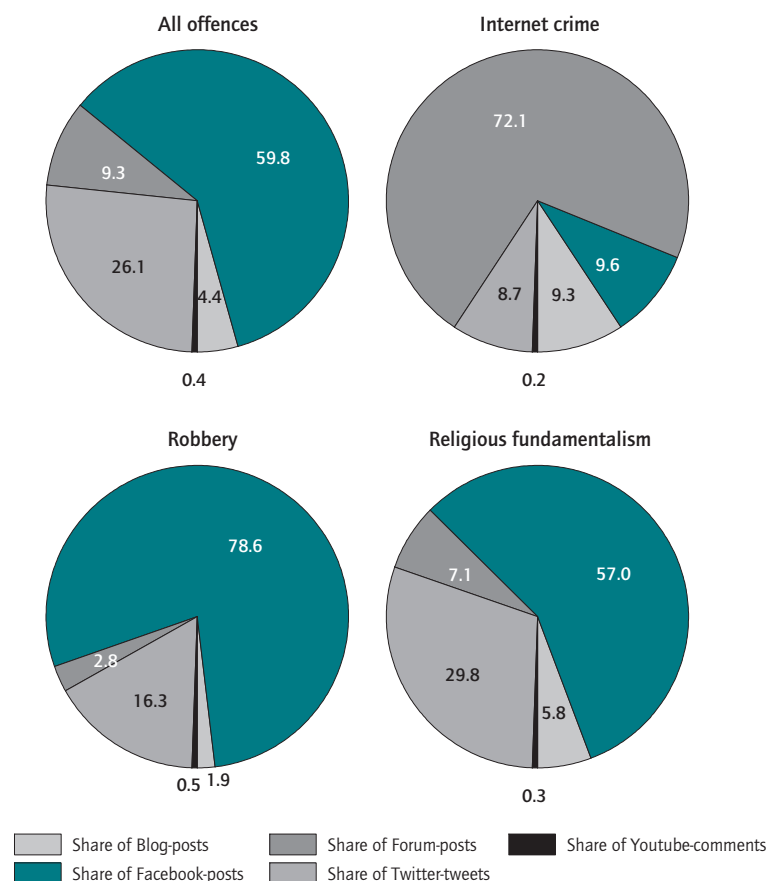
<sup>13</sup> Otherwise a distorted picture of densely populated urban areas would have been created. For example, more than 45,000 posts were assigned just to Berlin.

<sup>14</sup> The place where the crime is committed is usually far from where the resulting damage occurs.

<sup>15</sup> Mahrt and Scharnow, “The Value of Big Data in Digital Media Research,” 25.

Figure 1

### Distribution of social media sources for different types of offences in percent



Source: Bug, M.; Kroh, M.; Meier, K.; Rieckmann, J.; van Um, E.; Wald, N. (2015): WISIND-datasets: social media/Google Trends. Calculations by DIW Berlin.

© DIW Berlin 2015

The relative frequency of posts regarding different offences varies over sources: For instance, robbery is written about mostly on Facebook.

ny.<sup>16</sup> This is followed by tweets and posts on discussion forums and blogs (second to fourth place) with YouTube comments accounting for only a very small part of the sample. It is worth noting that different media platforms were associated with different categories of crime. The vast majority of posts on internet crime come from discussion forums in which users can get advice on how to protect themselves from internet crime or recover from any losses they have already incurred (see Figure 1). Posts on robbery are discussed and shared on Facebook with above-average frequency. Fundamentalism—considered to be one of

<sup>16</sup> See also K. Busemann and C. Gscheidle, "Dabei sein ist alles – zur Nutzung privater Communities. Ergebnisse der ZDF-Studie Community 2011," Media Perspektiven, no. 7–8 (2012): 380–390.

the most important issues in the period for which the sample was collected—is mentioned with equal frequency in blogs and discussion forums but more frequently on Twitter and Facebook.

### Comparing Social Media with Google Trends

How big is the difference between expressive and exploratory information behavior on the Internet? A Google Trends ranking of the frequency of certain search terms and a comparison with word frequencies analyzed in the sample of social media data for the period from June to October 2014 can shed some light on this (see Table 2). For each data source, the most frequent term was assigned a value of 100 and the other terms were assigned values in proportion to this maximum value. In Google Trends, queries containing the term "murder" clearly predominated; other terms ranked far lower. The rest of the ranking is similar: "theft," "burglary," and "bodily harm" rank near the very top, both in Google Trends data and on social media. Evidently Germans rarely searched for terms that were associated with religious fundamentalism (for example, "Islamist" or "Salafi") in this period—the most obvious difference between Google search and social media.

Table 2

### Comparison of ordinal rankings of frequencies in social media and search engines<sup>1</sup>

Scale from 1 to 100

Offence	Google Trends	Social Media
Homicides (murder)	100	50
Theft and burglary	26	100
Bodily harm	26	61
Drug related crimes (drugs)	18	37
Robbery	9	37
Internet crime (virus)	9	18
Sexual offences (rape)	9	15
Religious fundamentalism (Islamist/Salafi)	5	99
Organized crime (human trafficking)	2	7
Politically motivated crime (Right wing & Left wing extremism)	2	14

<sup>1</sup> June through October 2014

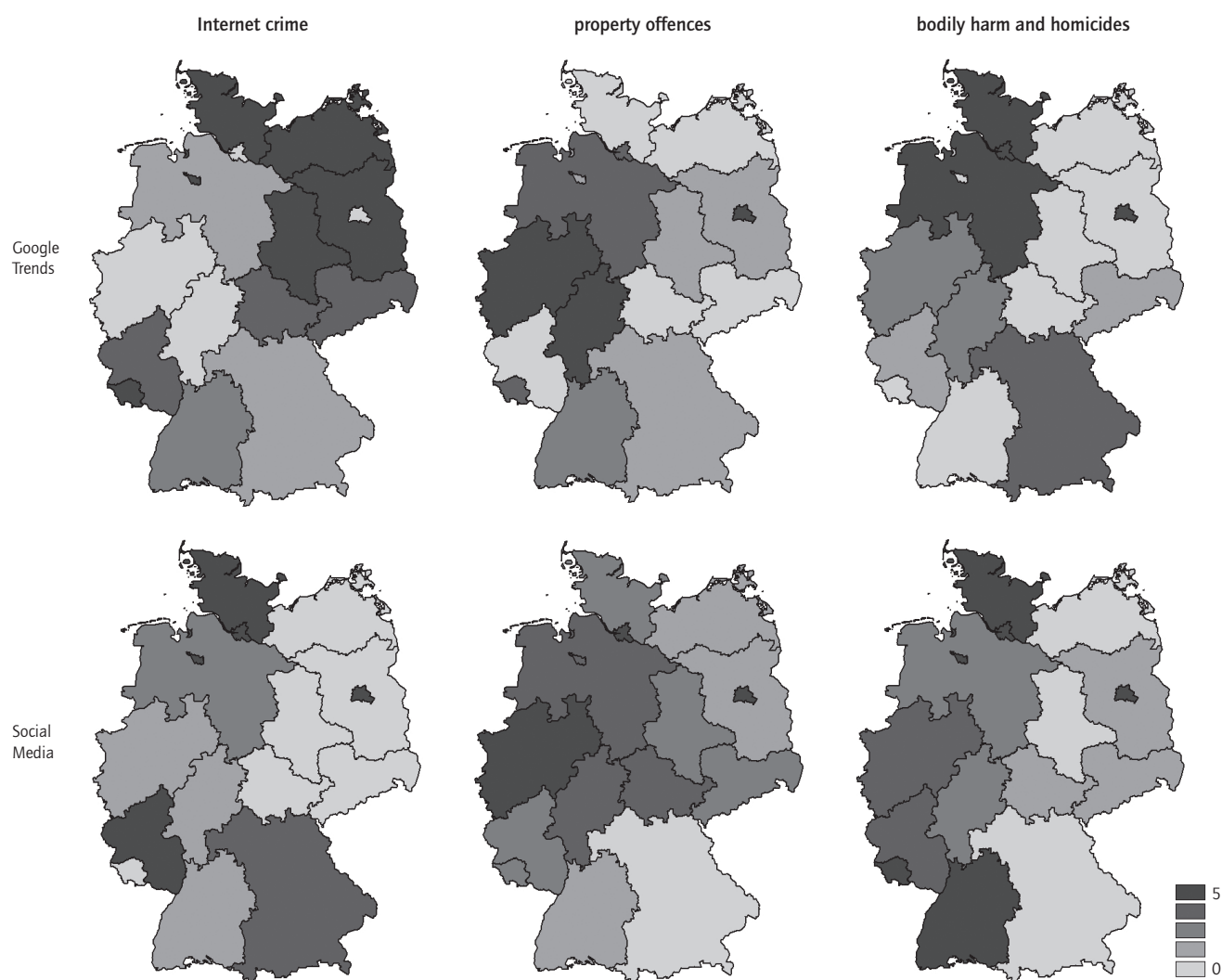
Source: Bug, M.; Kroh, M.; Meier, K.; Rieckmann, J.; van Um, E.; Wald, N. (2015): WISIND-datasets: social media/Google Trends. Calculations by DIW Berlin.

© DIW Berlin 2015

Terms related to theft, burglary and bodily harm are ranked quite high in Google Trends as well as in social media.



Figure 2

**Spatial distribution of Google Trends search queries (above) and posts in social media (below) in quintiles**

Source: Bug, M.; Kroh, M.; Meier, K.; Rieckmann, J.; van Um, E.; Wald, N. (2015): WISIND-datasets: social media/Google Trends. Calculations by DIW Berlin.

© DIW Berlin 2015

Terms related to Internet crime yield opposing results in Google Trends and social media: in East Germany Internet crime terms display higher frequencies in Google, in social media they are mostly mentioned in West Germany.

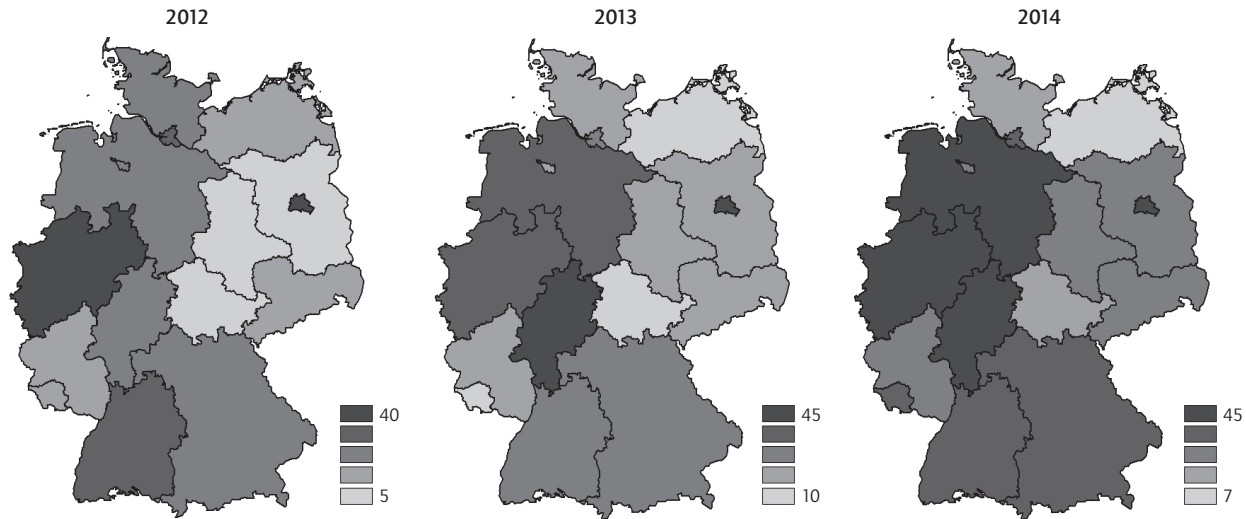
For a further systematic comparison of the two data sources, three classes of crime were formed: the first category contains various types of internet crime, the second consists of terms related to property crime, and the third comprises crimes involving death and physical injury. While Google Trends does provide data for regions below the level of federal states, these data are not complete and not available for every administrative district. For comparison purposes, therefore, the localized data from social networks were aggregated to represent federal states, and Google Trends data were obtained for the period from June to Octo-

ber 2014 (Figure 2 shows a comparison of geographic distributions).<sup>17</sup>

The contrast between social media and Google Trends regarding the class of internet crime is striking: Google data contains more searches from northern and eastern German states; the search volume for southern and

<sup>17</sup> A direct comparison of the colors used in the maps (Google Trends and social media) is not possible because the colors represent percentages of different maximum values. Meaningful inferences about the relative frequency of terms can therefore only be made for each of the maps separately.

Figure 3

**Google Trends search queries regarding property offences<sup>1</sup>**

<sup>1</sup> 2012 – 2014, equal intervals

Source: Bug, M.; Kroh, M.; Meier, K.; Rieckmann, J.; van Um, E.; Wald, N. (2015): WISIND-datasets: social media/Google Trends. Calculations by DIW Berlin.

© DIW Berlin 2015

The number of search queries concerning property crimes increased all over the country between 2012 and 2014 (from left to right), in most federal states.

western states is smaller in comparison. These results are similar to the fear of crime measurements made by the WISIND project, which found that northern Germany tended to have higher fear rates.<sup>18</sup> The opposite is true of expressive Internet use, i.e., the use of social media. While the data from social media must be treated with caution—for instance, it was possible to localize only five percent of the total posts on internet crime—this contrast presents an interesting topic for further research.

In the class of property crimes, the German states of Berlin, Hesse, and North Rhine-Westphalia along with Lower Saxony are at the top of the Google Trends ranking with high search volumes for terms like “burglary,” “theft,” and “alarm system.” In social media there is a slightly more distinct north-south divide, with at the same time less obvious differences between West and East Germany.

Crimes involving death and physical violence show both similarities and differences between data sources and regions. In the north-western German federal states (with the exception of Bremen) as well as in Berlin posts about this class of crimes are frequent. Search queries

concerning the respective crime field are less frequent in Baden-Wuerttemberg, Bremen and Saarland than in Bavaria, the exact opposite holds true for social media.

### Comparing Data with Actual Crime Rates and Fear of Crime

Google Trends makes it possible to select a time period and compare data spanning several years. Annual averages were calculated on the basis of weekly or monthly data for each of the German states.<sup>19</sup> Between 2012 and 2014, searches related to property crime tended to rise in most states (see Figure 3). Every year, Google users—with the exception of those in Mecklenburg-Western Pomerania and Schleswig-Holstein—searched for terms related to the topic of property crime more frequently than in the previous year, although the increase has tailed off across Germany since 2014.

<sup>18</sup> See also the article by M. Bug, M. Kroh, and K. Meier in this issue of DIW Economic Bulletin.

<sup>19</sup> Seasonal variation in the occurrence of the search term “burglary” would be another interesting topic of study, but one which cannot be examined more closely here. Limiting the data sample from social media to the period between June and October 2014 does not allow any reliable conclusions to be drawn about seasonal changes in the attitudes and interests of Internet users.



The very limited suitability of search engine data as an indicator of objective crime rates—not fear of crime—becomes evident when they are compared to PKS data<sup>20</sup>. PKS reports for the years 2012 and 2013 show only a slight rise of about 0.1 percent in the number of burglaries and thefts in Germany.<sup>21</sup> Based on these figures, it would seem that Google Trends cannot be used to make any direct inferences about actual crime rates in Germany. At the federal state level, however, Google Trends correctly reflected developments in crime rates 11 times in the period from 2012 to 2013. It is highly doubtful, however, that these figures can be used to form a reliable overall picture, much less to make predictions.

A corresponding comparison of trends in the data from social media and the PKS cannot be made because of the data collection period, but here too differences can be seen, specifically differences between the number of posts and the number of cases associated with particular locations.

To assess whether the findings presented here from social media and search engines are suitable for use in mapping fear of crime (as opposed to objective crime rates), a comparison with representative survey data is required. A survey was carried out by DIW Berlin's WISIND project; it polled 12,000 people in Germany regarding their concerns about becoming the victim of various crimes.

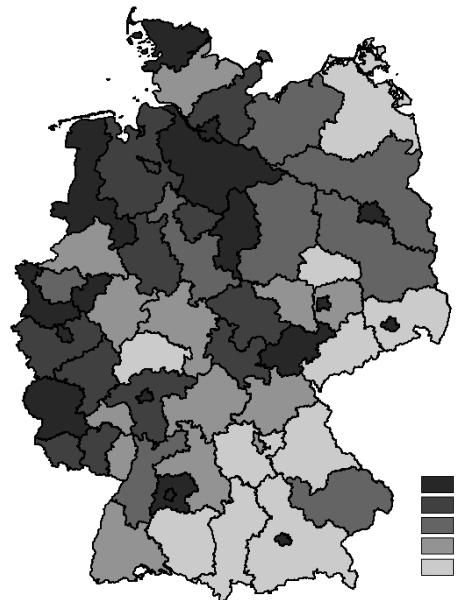
A comparison with the data measuring fear of crime<sup>22</sup> in 2014 (presented in another article in this issue of DIW Economic Bulletin) shows that some of the geographic patterns associated with the ten crime categories are reflected in the regional distribution of the crime indicator “subjective fear” (see Table 4).<sup>23</sup> This is particularly true of Schleswig-Holstein, Berlin, the northern part of Lower Saxony, and large parts of North Rhine-Westphalia. The districts around Stuttgart<sup>24</sup> and the administrative regions of Karlsruhe and Freiburg can be seen in both maps, both of which show the same levels of Internet activity and fear of crime.

At the same time, the comparison reveals several differences at the regional level. At first glance, these might be attributable to the differing data collection periods; a fur-

Figure 4

### Post frequencies in social media for all ten offence groups<sup>1</sup>

per 100,000 inhabitants in quintiles on regional level



<sup>1</sup> June through October 2014

Source: Bug, M.; Kroh, M.; Meier, K.; Rieckmann, J.; van Um, E.; Wald, N. (2015): WISIND-datasets: social media/Google Trends. Calculations by DIW Berlin.

© DIW Berlin 2015

The spatial distribution of posts within all ten offence groups displays similarities with the mapping of the WISIND crime fear indicator on regional level.

ther distinction can be seen between the simple mention of issues related to crime, on the one hand, and specific questioning about concerns regarding various forms of crime on the other.<sup>25</sup> In some cases, this results in regions being depicted on the maps in contrasting colors. The region of Trier shows a relatively low level of subjective fear of crime, i.e., people there tend to worry less that they will become victims of crime. On social media, however, this region is among those with the highest numbers of posts relating to crime. The same contradiction can be observed for Middle Franconia, Upper Palatinate and Lower Bavaria. Overall, the data collected from the Internet clearly reflect fear of crime much more accurately than actual crime rates, and they could be a good indication of differences in attitudes between regional populations.

<sup>20</sup> This statement assumes there is no significant change in the number of unreported crimes.

<sup>21</sup> Federal Criminal Police Office, Jahrbuch Polizeiliche Kriminalstatistik (2012-2014).

<sup>22</sup> For a graphic showing crime indicators of subjective fear, see also the above-mentioned article by Bug, Kroh, and Meier.

<sup>23</sup> The graphic is based on a rendering of regional social media activity in quintiles.

<sup>24</sup> It consists of the Rems-Murr, Böblingen, Esslingen, and Ludwigsburg districts.

<sup>25</sup> Media-use behavior, influenced by demographic factors, can also play a key role in regional differences.

The ability to regionalize the social media data sample widens its potential application beyond simple comparisons with other figures. The prominent position in the data of the Braunschweig-Salzgitter-Wolfsburg region is striking: it has the maximum value by a large margin for frequency of terms related to religious fundamentalism. This was the case even months before media coverage of the arrests of young Islamic extremists in Wolfsburg and other coverage of the carnival parade in Braunschweig that was recently canceled due to direct threats of a terrorist attack by Islamic extremists.<sup>26</sup>

## Conclusion

The findings of the analysis of data from social media and search engines indicate that, because of their intrinsic qualities and also for systemic reasons, they are not suitable for creating a reliable picture of real regional crime rates. However, the data certainly can be used to form a picture of subjective perceptions of

crime in regional populations and as such can function as a cost-effective data source supplementing traditional surveys on fear of crime. The analysis of social media content shows three limitations to interpreting the data as an indicator of actual crime risk. First, social media are heavily influenced by media effects and episodes of heightened interest in certain topics; this typically takes the form of retweeting, reposting, and sharing excerpts from other media. Second, it is not easy to localize the data. Data can be localized only when users disclose this information.<sup>27</sup> Their willingness to do this, however, depends on the context of the posts, with the result that some terms related to specific crimes can be localized more easily than other terms. Third, the selection of Internet platforms is very important, as there are considerable variations in the types of media used to discuss different categories of crimes. Comparisons with actual crime rates are further limited by a sample bias in favor of younger users, who are much more active on social media.

<sup>26</sup> The map depicting the frequency of terms related to the "religious fundamentalism" crime category is not included in this DIW Economic Bulletin article.

<sup>27</sup> A further challenge is the use of place name spellings that deviate from official orthography, as the matching process cannot be fully automated and is time-consuming.

Johannes Rieckmann is Research Associate within the Department of Development and Security at DIW Berlin

Jan-Lucas Schanze is Student Assistant within the Department of Development and Security at DIW Berlin

JEL: H79, K42, R12, Z18

**Keywords:** Crime, crime statistics, indicator, perception, security, social media, social network, search engine



DIW Berlin – Deutsches Institut  
für Wirtschaftsforschung e. V.  
Mohrenstraße 58, 10117 Berlin  
T +49 30 897 89 -0  
F +49 30 897 89 -200

#### Publishers

Prof. Dr. Pio Baake  
Prof. Dr. Tomaso Duso  
Dr. Ferdinand Fichtner  
Prof. Marcel Fratzscher, Ph.D.  
Prof. Dr. Peter Haan  
Prof. Dr. Claudia Kemfert  
Dr. Kati Krähnert  
Prof. Karsten Neuhoff, Ph.D.  
Dr. Kati Schindler  
Prof. Dr. Jürgen Schupp  
Prof. Dr. C. Katharina Spieß  
Prof. Dr. Gert G. Wagner

#### Reviewer

Dr. Anika Rasner  
Dr. David Richter

#### Editors in chief

Sabine Fiedler  
Dr. Kurt Geppert

#### Editorial staff

Renate Bogdanovic  
Andreas Harasser  
Sebastian Kollmann  
Dr. Claudia Lambert  
Dr. WolfPeter Schill

#### Translation

HLTW Übersetzungen GbR  
team@hlw.de

#### Layout and Composition

eScriptum GmbH & Co KG, Berlin

#### Press office

Renate Bogdanovic  
Tel. +49-30-89789-249  
presse@diw.de

#### Sale and distribution

DIW Berlin

Reprint and further distribution – including extracts – with complete reference and consignment of a specimen copy to DIW Berlin's Communication Department (kundenservice@diw.berlin) only.  
Printed on 100 % recycled paper.