SPECIAL ASPECTS OF DIGITAL INFORMATION ACQUISITION ON THE INTERNET USED IN CRIMINAL INVESTIGATIONS AND PERSONAL DATA ANALYSIS

ОСОБЕННОСТИ СБОРА ЦИФРОВОЙ ИНФОРМАЦИИ В СЕТИ ИНТЕРНЕТ ДЛЯ ЦЕЛЕЙ РАССЛЕДОВАНИЯ ПРЕСТУПЛЕНИЙ И АНАЛИЗА ДАННЫХ О ЛИЧНОСТИ

ASPECTOS ESPECIALES DE LA ADQUISICIÓN DE INFORMACIÓN DIGITAL EN INTERNET UTILIZADA EN INVESTIGACIONES PENALES Y ANÁLISIS DE DATOS PERSONALES

Recibido: 1 de abril de 2019. Aceptado: 15 de mayo de 2019

Written by:
Nina V. OLINDER
https://orcid.org/0000-0003-1489-9574
Vera V. ROMANOVA
https://orcid.org/0000-0002-9901-7667
Ilmira Z. TOVYSHEVA
https://orcid.org/0000-0001-6263-0065
Stanislav V. YUNOSHEV
https://orcid.org/0000-0002-3537-7460
Evgeniya A. GAMBAROVA
https://orcid.org/0000-0002-9984-2819

Abstract

The article describes the course and results of the social experiment, which was conducted on the basis of the scientific and educational laboratory of criminology and forensic examinations of the Samara University (2017/2018 academic year) and the criministics laboratory of the Department of Criminal Law and Process of Togliatti State University (2018/2019 academic year). The purpose of the experiment was to compare the effectiveness of various ways of obtaining information about a person (oral survey, gathering information on the Internet) to select the method of collecting information that is optimal in time and labor. For this, it was proposed to different groups of participants in the experiment to collect information about certain individuals according to the set parameters by conducting a traditional survey and by analyzing profiles in social networks. As a result of the study, the difficulties of the prospects for using social networks in forensic science are identified.

Аннотация


156 PhD in Law, associate professor, the Head of Constitutional and Administrative Law Department, Togliatti State University
157 PhD in Law, associate professor of Constitutional and Administrative Law Department, Togliatti State University
158 PhD in Economics, associate professor, Department of Government, Bashkir State University
159 PhD in Law, associate professor, the Head of Criminal Law and Procedure Department, Togliatti State University
160 ООО Pravovoye pole, in-house counsel, Department of Criminal Procedure and Forensic Science, Samara National Research University (Samara University)
a range of information that users prefer not to share about themselves has been identified, a work method has been developed for law enforcement officials of the CIS countries with social networks in Runet space.

**Keywords:** social networking, security and privacy, crime investigation, cyber-investigation, personality, digital evidence.

**Resumen**

El artículo describe el curso y los resultados del experimento social, que se realizó sobre la base del laboratorio científico y educativo de criminología y exámenes forenses de la Universidad de Samara (año académico 2017/2018) y el laboratorio de criminalística del Departamento de Derecho Penal. y Proceso de la Universidad Estatal Togliatti (año académico 2018/2019). El propósito del experimento fue comparar la efectividad de varias formas de obtener información sobre una persona (encuesta oral, recopilación de información en Internet) para seleccionar el método de recopilación de información que sea óptimo en tiempo y trabajo. Para esto, se propuso a diferentes grupos de participantes en el experimento recopilar información sobre ciertos individuos según los parámetros establecidos mediante una encuesta tradicional y el análisis de perfiles en las redes sociales. Como resultado del estudio, se identificaron las dificultades de las perspectivas para el uso de las redes sociales en la ciencia forense, se identificó una gama de información que los usuarios prefieren no compartir sobre sí mismos, se ha desarrollado un método de trabajo para los funcionarios encargados de hacer cumplir la ley de Países de la CEI con redes sociales en el espacio Runet.

**Palabras clave:** redes sociales, seguridad y privacidad, investigación de delitos, investigación cibernética, personalidad, evidencia digital

**Introduction**

The contemporary public approach to the ways and forms of communication, information sharing, etc. is notably changing. This is partly driven by the development of the worldwide net and expansion of virtual relations. Instant messaging and social network communication have kicked into high gear (Gambarova, E.A., Olinder, N.V., 2017). Increased popularity of social media (Social media (2018)) (social networks, messengers, online computer games, etc.), widespread acceptance of virtual databases, online banking, cloud storage, and other tools that facilitate easy and instant communication and receiving / offering services necessitates statutory regulation of such matters, as well as induces elaboration of a new virtual space methodology in forensic science (Forensic literature study and scientific & practical conferences review have showed that media hits of ‘virtual’, ‘digital’, ‘electronic’ concepts, etc. have increased by 57%).

Virtualization of social interactions manifests itself in a significant increase in the number of social network users, their social media activity – both through instant messaging, and getting various photos and comments about themselves, their loved ones, their interests, as well as thoughts similar to theirs, posted online – on their webpages in social networks.

Social media are most commonly defined as a form of mass communication via the Internet or as a combination of web-based and mobile technologies to help users to communicate with one another (correspondence, networking, exchange of information, news, etc.).

One of the Internet features is that once uploaded to social media the data accumulates and stores there for a sufficient period. This very property can form the basis for such data processing by an investigator, interrogating officer or a detective. Since the information is instrumental in the cognitive activity of an investigator, finding the ways of a quicker and more complete
information acquisition (including such from social media) constitutes a priority in forensic science (Meshcheryakov, V.A., 2013, p. 87–92; Meshcheryakov, V.A., 2014). In various nations (e.g. the USA), social media are being extensively used in criminal investigations, and are considered as a valuable tool to search and find information about events and people of interest (Investigations (2017).

In Russia, using social media in criminal investigations and crime prevention has not become a common practice. Digital data is generally used by special units of investigative and inquiry bodies, however, a cut procedure for working with social media is still under development. Yet, on a practical level, digital data is brought into use by both investigative bodies (via special units), and the defense. For instance, in 2017, the city of Samara, a defender got an acquittal by using digital data from social media. A witness is one of the participants of the crime investigation procedure. Having studied the social media profiles of the witnesses, the defense attorney found out that the witnesses in the case (in which he was engaged as a defender of the suspect) had known each other; in fact, the pictures showed that some of these witnesses were wearing their police uniforms. Driven by aroused suspicion, the defense attorney filed a motion to the trial court requesting the social network owner to provide the correspondence between the witnesses. This correspondence served as a data source and contained the witnesses’ discussion about planting drugs on the attorney’s client and the ethical component of this case.

Therefore, using the social media information the attorney was able to initiate reception of evidence, which thereafter resulted in an acquittal. More interestingly, the attorney used social media in his work without having any vocational skills or knowledge in computer engineering, i.e. he was able to learn information outright, with no middleman – a computer specialist – involved.

Scientists in the field of criminal procedural law are increasingly raising the issue of “digital evidence”, “digital evidence base”, comes the understanding that the investigative practice cannot do without the collection and use of digital information (Chernyshov, V.N., Loskutova E.S., 2017, p. 199–203; Fatyanov, A.A., Doroshkov, V.V., 2017). This issue is given sufficient attention and scientists from foreign countries (Donald, A. 2013; Orin, S., 2018; Loia, V., Mattiucci, M., Senatore, S., Veniero, M., 2009).

The experience of foreign countries suggests that the use of digital evidence (information transmitted through computers and other hardware) plays an important role in investigating various types of crimes, including murder, rape, abduction, child abuse, harassment of minors, child pornography, fraud, theft, drug trafficking, computer intrusion, espionage and terrorism. However, until now there are problems in collecting and using such information (Casey, E., 2014).

The need for using social media information in criminal investigation is also indicated by Yuri Kamenetsky, PhD in Law, an Investigator for Major Cases of Belarus Investigative Committee. In his dissertation ‘Methods of Initial Investigation of Misappropriation by Official Misconduct in the Government Sector’ he writes that “prevention of counteraction by a plunderer and his/her corrupt relationships is directly linked to the analytical efforts aimed at determination of the plunderer’s social and family ties, and taking preventive measures. Along with the common searching methods used by an investigator, monitoring of social media publicly available in the World Wide Web, e.g. VKontakte, Odnoklassniki, Facebook, etc. is considered one of the most effective investigative techniques. A study conducted by means of searching the Internet has shown that 31.1% of a plunderer’s personal data is currently available in the public domain in social media profiles. This operative search activity seems promising” (Kamenetsky, Yuri. F., 2016). Therefore, monitoring of social media data available in the public domain provides far beyond than operational detection of a person’s social connections, material and marital status, but also identify some basic data about the individual (Colombini, C., Colella, A., 2011, p.).

This information captured in the photos (friendly relations, recreations, availability of property, vehicles, and life style) facilitates in setting out leads on the financial standing and material wealth, leivable property, etc.

Methods

In the course of the experiment and during preparation of this paper, the following methods were used. General scientific methods: observation, analysis and synthesis. Observation refers to a targeted collection of information about the facts of a person’s behavior and activity,
in various natural conditions. In this experiment, information was collected about a person’s identity on the Internet (or by interviewing – the first stage of the experiment); also, collected was the information about the actions of persons who collected such information. The performed experiment has met all the necessary requirements for observation as a method of scientific research:

- purposefulness – the presence of a clear target setting: to collect information about particular people;
- consistency – the presence of unchanged ‘objects’ of research: the people, whose personal data had to be collected; data acquisition continued for two years, and was implemented by different groups of students; the conditions were both repetitive (‘beacons’ for data collection) and changing: inquiries and receiving data on the Internet, data acquisition only on the Internet, only on social media, from any Internet sources;
- naturally occurring process – on the one hand, the students participating in the experiment performed the task within the framework of the “Forensic Science” discipline; on the other hand, the “objects of research” social media profiles – were not specifically created for the experiment, people continued to maintain their social media pages, uploaded information, posts, photos, updated their profiles, etc.

Analysis as a method of scientific research was used in ‘processing’ the information received – analysis of the social media page contents, systematization of the information received, and compilation of tables. Analysis helped to identify information attributes and data groups. Furthermore, analysis refers to the main method of an investigator’s cognitive activity; it is the analysis that facilitates an investigator or an interrogating officer in reviling consistent patterns and defining important evidence.

Synthesis (from the Greek ‘Synthesis’ – connection, combination, compilation) refers to the connection of various elements, the aspects of an object into a single whole (the system). Only when using synthesis as a method of scientific research it is possible to get a ‘complete’ picture of a personality by comparing all the data received about it, rather than being based, for example, on data obtained from a page in one social network only.

In addition to general scientific methods, a comparative legal method was used – the best foreign practice to using information about a person from the Internet in criminal investigations has been studied, the basic consistency of the ‘digital data’ institution development has been revealed. Unlike the general scientific approaches, this method application to address the problem allows us to reflect the specifics of collecting information about an individual on the Internet. Furthermore, this research method can be used to assess the impact of the court practice on the development trends of legal regulation of information circulation.

Since the circulation of information on the Internet (including personal data) is interdisciplinary in its nature, and is regulated by the legal provisions of several branches of law (civil law, civil procedural law, criminal law, criminal practice, data protection law, administrative law, etc.), applied was an interdisciplinary legal research method, which allowed a comprehensive analysis of legal provisions and evaluation of their impact on the circulation of information on the Internet.

**Results and Discussion**

All of the aforesaid speaks of the need to develop procedure for efficiently obtaining and processing information about an individual on the Internet. For this, over a period of two years, the authors conducted an experiment to develop skills for collecting information on the Internet (through social media) – completeness, accuracy and efficiency of obtaining information about a person on the Internet were developed. The experiment was conducted at Samara National Research University (2017/2018 academic year) and Togliatti State University (2018/2019 academic year). As part of the experiment, an attempt was made to develop an algorithm for searching and collecting public information about a person based on information acquisition on the Internet.

The purpose of the experiment is to compare the effectiveness of various ways of obtaining information about a person in order to find a time-and-labor optimal way to collect information.

Experimental methods: information collection, information synthesis, data analysis, observation, questioning, interviewing, information compilation, construction of charts and tables (graphical method), and programming.
The experiment was attended by third- and fourth-year students of these universities, who have already obtained general knowledge as part of the educational program: the court and law enforcement agencies, criminal law, criminal procedure, documentation management. The students were divided into subgroups; each group was asked to collect information about certain individuals according to specified parameters by conducting a traditional survey and by analyzing social media profiles. The research part of the experiment was implemented in several stages.

At the first stage, the task was simplified: to collect information about the subject (without specifying it). The task was to compare the volume and accuracy of information obtained over social media, and those of information obtained through interviewing. The students were divided into three groups (6 people in each group), each of which was assigned to collect information about a particular person through social media and by means of oral interviewing. Prior to implementation, the individuals whose information had to be collected had given their consent to participate in the experiment. For convenience of description, let's call them subject 1, subject 2, and subject 3.

Subject 1: 29 years old, has been using the Internet for more than 10 years, social media – for eight years; has had over five accounts in various social media (some of them with limited access); has been a member of various forums, has had several scientific publications and has been mentioned in the media several times.

Subject 2: 40 years old, has been using the Internet for more than 10 years, social media – for eight years; has had two social media accounts; has been a member of professional forums, has had over 60 scientific publications.

Subject 3: 47 years old, is not an active user of the Internet, does not have any social media accounts.

Each group of students was divided into two subgroups (three people in each subgroup): the first subgroup was collecting information by interviewing, while the second group was making use of social media only. The experimental participants had to describe in detail their information acquisition technique, as well as record the time spent on the task implementation. The task deliverables is the ‘portfolio’ of the subject under study. No sharing the results of interviews or the results of social media processing was allowed, nor was it allowed to use special search applications to acquire information over social media. The task had to be completed in 7 (seven) days.

The first experimental phase has delivered the following results. The time spent on searching for information about a person by means of interviewing averaged three working days. It took an average of two and a half hours to search for information about a person on the Internet.

The completeness of the portfolio also turned out to be different. Thus, the experimental participants who conducted the survey have found far less information on active Internet users than those who have analyzed social media. In particular, the experimental participants who searched for information on the Internet managed to learn about the person’s bad habits, most frequently visited places, surrounding community, and to learn enough about the type of activity, financial situation, etc.

However, more information about people who are inactive users of the Internet could be obtained by means of interviewing them. At the same time, this information differed from the information obtained by the groups that worked on the Internet. Despite the fact that the third participant in the experiment was not a social media user, the students were able to find out information about his parents and the type of their activity, the cars owned, the absence of registration as an individual entrepreneur and other information of this kind on the Internet.

Thus, the first experimental stage has supported the assumption that using the Internet facilitates in obtaining more information in less time, but this rule applies only to individuals who are registered on social media or actively use the Internet, or are being members of professional communities, etc.

As a result of multiple repetition of the experiment in its various modifications, it has been revealed that it takes an average of 2.5 hours to compile the most detailed portfolio about an individual (including the time required for this information searching and processing). Speaking about information processing, we mean that the information uploaded by users on social media is not always reliable and in order to determine its preliminary reliability, the researcher has to spend a certain amount of time. At the very beginning of the article, we identified the initial goal, which is to find an optimal data processing technique, so it was decided to conduct the
second stage of the experiment aimed at finding automated data processing techniques. The idea behind this experiment is that students (115 people – the experimental participants) were suggested to find information about certain people (3 people) according to specified parameters (place of residence, place of study, joining founding members, existence of property, debts and fines, participation in lawsuits, travelling and business trips, leisure, family members, close friends).

It was critical that the experimental participants did not use special software (search programs) and searched for open source information only.

The goal of the second experimental stage was to find out what information about an individual could be obtained from open sources without applying special knowledge or software. The experiment has demonstrated that one can easily find information about the city of residence (67% of the participants managed to find this info), although only 5.7% of the participants could define a specific address. 79% of the participants were able to identify places of work and study. Information about membership in founders, shareholders, individual entrepreneur status, etc. was found by 22% of the experimental participants, however, only 10% of the participants succeeded in finding information about existence of property, though this data was insufficient. Less than 2% of participants could detect debts, fines and loans; less than 1% of participants found out that a person participated (as one of the parties) in lawsuits; more than half of the participants (50.6%) could find information about travelling and business trips; 29.5% of participants found information about parents, 52.5% of participants succeeded in finding info about spouses; info about siblings was acquired by 19% of participants, and about 20.9% could find information about friends. Here we are talking only about sufficient and reliable information. Data searching and processing (including determining its reliability) took approximately 3.5 hours.

Thus, certain parameters (‘beacons’) that are freely available on social media and that can be available to any user – address, place of work, travelling (both business trips and recreation), immediate environment, hobbies and interests – have been identified during the second experimental stage. This open access information can be used in criminal investigations, e.g. when collecting information about possible participants of an organized crime group, when preparing for individual investigative actions (for example, interrogation, confrontation) or in general, when planning an investigation of certain types of crimes.

**Findings**

Thus, we can sum up the following results of the experiment.

1. ‘Personal interviewing’ vs. ‘Searching for information on the Internet’ data acquisition technique

<table>
<thead>
<tr>
<th>Personal interviewing</th>
<th>Searching for information on the Internet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Good points</strong></td>
<td><strong>Drawbacks</strong></td>
</tr>
<tr>
<td>1) One can collect</td>
<td>1) Always the subjective opinion of the person being interviewed;</td>
</tr>
<tr>
<td>information about the</td>
<td>2) 100% effective only for those who do not use the Internet;</td>
</tr>
<tr>
<td>person’s character,</td>
<td>3) The amount of information depends on who the respondent is in</td>
</tr>
<tr>
<td>his habits,</td>
<td></td>
</tr>
<tr>
<td>environment, hobbies,</td>
<td></td>
</tr>
<tr>
<td>find out the age,</td>
<td></td>
</tr>
<tr>
<td>scope of activities,</td>
<td></td>
</tr>
<tr>
<td>and so on;</td>
<td></td>
</tr>
<tr>
<td>2) With special</td>
<td></td>
</tr>
<tr>
<td>knowledge (psychological), it is possible to determine whether there is a connection between the</td>
<td></td>
</tr>
<tr>
<td><strong>Drawbacks</strong></td>
<td><strong>Good points</strong></td>
</tr>
<tr>
<td>1) Time spent to find information (2-3 hours);</td>
<td>1) A person is not a social media user;</td>
</tr>
<tr>
<td>2) The amount of</td>
<td>2) Private accounts;</td>
</tr>
<tr>
<td>information collected</td>
<td>3) False information that a person uploads about himself/herself</td>
</tr>
<tr>
<td>exceeds the amount</td>
<td></td>
</tr>
<tr>
<td>collected during an</td>
<td></td>
</tr>
<tr>
<td>interview (for active Internet users);</td>
<td></td>
</tr>
</tbody>
</table>
person being interviewed and the person about whom the information is being collected.

3) You can find personal information not only on social media, where the person uploads it voluntary, but also from other information resources (cadastral records, fines database, information on participation in organizations, events, etc.);

4) You can track connections between people, check communities, identify groups.

5) Time spent on interviewing (2-3 days);

6) For more sufficient information, it is necessary to interview more people (2-5 people) from different spheres of a person’s life.

The data retrieval procedure is as follows:

1. Use a search engine or a retrieval engine installed on the subject’s computer (the device must be connected to the Internet). Enter the information already known about the object (a person’s surname, first name, year of birth, place of work, etc.) into the search box.

2. Generally, the first few links may refer to an employee profile on the corporate website, including links to social media accounts (e.g. VK, Facebook, Instagram), or other profiles where the object under study has been checked. Any of these resources may contain personal data (a person’s name, surname, age, date of birth, marital status, information about family members, employer and current position, education profile, place of birth, etc.) – this information is available in the profile header.

3. Capture the acquired information, analyze it and draw initial conclusions (age, education, place of living, marital status, etc.).

4. Analyze the photos published on the social media accounts to obtain information about

2. The experiment revealed a scope of information that the subjects do not always upload on social media (frequently visited Internet pages, subject search requests, etc.), searching for which requires additional special programs similar to web marketing tools collecting information about a person that is used to target ads.

3. As a result of the experiment, an algorithm was developed for collecting information about a person on the Internet, which can be used in criminal investigating.

Algorithm of personal data acquisition in the Internet.

Object of research: web pages of VKontakte (otherwise known as VK), Instagram, and Facebook social media, as well as the others who are being collected information about.

Subject of research: an investigator, an interrogating officer, a detective.

Research techniques: data acquisition, composition, observation, comparison, and analysis.
private, friendly, and corporate connections. This information can be confirmed by checking a friend / subscriber lists, analyzing the profile comments, etc.
5. Special attention should be paid to the 'official information', which is to be captured by taking a screenshot of document images with ID attributes (e.g. passport, certificate, driving license), a vehicle registration number, cadastral numbers to real estate properties, etc.
6. Analysis of groups, communities, and events helps identify important characteristics of a psychological profile: hobbies, interests, social attitude, religious and political beliefs.
7. Photos may become a source of information about person’s material level, income level, possible trips abroad or to other cities. They can also facilitate in identification of the object location at a certain time (notably, if the object posts geo-referenced photos, adds location or hashtag to a picture, etc.)
8. If the information on the object’s page is insufficient, it is advisable to check the webpages of the object’s relations (if any) and their friends.
9. Users of several profiles often link their accounts (pages) on various social media or provide address references to other profiles on one social media platform. This provides the subject of research extra means to find information about the object of research.
10. A generalized analysis of the entire information acquired on social media.
11. Reenter the object’s surname and name and then check the information in the professional communities, references in the media, publications, participation in events, etc. For example, a person is a member of a charitable organization, attends certain events, or due to his/her professional activities, gives interviews, participates in conferences, etc.
12. Check by key elements (name, date of birth, city) on the official portals (e.g. on the state cadaster site, by entering the address you can check the ownership of the land, by entering the system of fines, you can check if the object owns a car, etc. It depends on a particular task of collecting information).

While data searching and processing, it is advisable to take notes, draw charts, sketches in order to get a more complete picture.

In the course of cognitive activity on data assessment and analysis, the presented technique allows the researcher to establish the facts of comparison of information. If a psychologist is involved in the process of analyzing information about a person acquired on the Internet and social media, a psychological portrait of a person can be made (the third stage of the experiment has not been completed yet).

Notes. The experiment has not been completed yet; its third stage is currently being conducted – the study of a psycho-type of a person based on the information posted on social media. Summing up the two stages of the experiment, we can conclude that searching for information about an individual on the Internet is less labor-intensive than collecting information by means of interviewing. It is possible to simultaneously search and process information about several people on the Internet; when collecting and analyzing information on the Internet, not only ‘textual, character-based’ information can be collected, but also photo and video materials posted on the Internet can be used.

References
Fatyanov, A.A., Doroshkov, V.V. (2017). On the definition of "computer information" in the Russian criminal law // Information law, No. 3. (in Russ.)
Investigations (2017) [Electronic source]. Retrieved from


