

# SOCIOLOGICAL RESEARCH METHOD AND DIGITAL TRACES OF BIG DATA IN ECONOMIC SYSTEMS

Anastasia D. Arefyeva, Deerfield Academy

## ABSTRACT

**Aim of the study:** *Digital transformation has affected almost all sectors of the world economy, increasingly and consistently penetrating into information technology, forming a digital environment. At the same time, sociology, which is based on data collection and processing by various population groups, service types, areas of life and other parameters, is no exception.*

**Methodology:** *With the growing social digitalization, social media data the volume of which has increased significantly are becoming more and more popular, having concentrated a huge number of participants who give their opinions in networks taking part in public voting on certain issues, thereby leaving the so-called "digital trace." It is the given concept that underlies "Big Data" that are used in sociology enabling to more reliably identify public trends in a particular area. Big Data technology is used in research of virtually any parameters and categories concerning major problems, including health care, education, employment, social sphere, manufacturing and industry, public service delivery, politics and others. Using Big Data simplifies data processing, and makes it more transparent, accurate and reliable as well. Despite the effectiveness of these digital categories, it should be noted that they lose efficiency with no human labor, since they are managed by the research activity.*

**Conclusion:** *The article is aimed to identify what new technologies are capable of when conducting sociological research using "Digital traces - Big Data," which is based on the Big Data use. The author concludes that it is Big Data that allows for a better and deeper study in sociology, which has the most reliable results. The theoretical analysis showed how new technologies have been integrated into the sociological research system, while determining their advantages and disadvantages. A sociological labor market study shows how Big Data can be used in sociology. It was concluded that the new technologies enable determining a number of qualitative characteristics of the studied phenomenon, while processing a large array of data that without Big Data cannot be processed.*

**Keywords:** Sociological Research, Digitalization, Big Data, Social Networks, Labor Market, Sampling, Competencies.

## INTRODUCTION

When using the Big Data component in sociology, human behavior rating methods are refocused with a greater emphasis on person's psychological temperament, causes of effects and behavior in a given context. At the same time, different simulation modeling options, economic and mathematical models, methods of mathematical modeling are used. However, social science at large has not developed immediately. The first attempts to employ Big Data were traced in the works created by (D. Boyd & K. Crawford, 2009) which were used to solve the problem of representativeness (Boyd D & Crawford K, 2009) Critical Questions for Big Data: Provocations

for a cultural, technological, and scholarly phenomenon. Information, Communication & Society 2012). The authors believed that it was Big Data in the context of both science and security that made it possible to obtain new information about society.

The authors suggested studying human behavior from the formalization perspective. However, the method was rather time-consuming and not widely applicable as a result of the large amount of data analyzed. In addition, the researcher should always be aware of what is given in the following text and as a result, the trend hasn't assimilated in sociology. At the same time, it proceeded modifying and a method devoid of these shortcomings was formed at the next stage of development. This method was suggested by another scientists A. De et al. (2015), which was based on the academic paper analysis edited regarding Mayer-Schonberger's remarks (Mayer-Schonberger & Cukier, 2014).

The A. De Mauro's approach was explicit, which implied that all data analysis is carried out using an automatic approach, enabling working with large texts of data, as well as process the text according to the prepared scheme. In this method, the text was encoded by using different categories which were ranked in accordance with the topics compiled by the algorithm and initially established by a human, as a result of which the method turned out to be inductive.

As early as in 2014, more advanced fields of sociology were developed through the use of Big Data, when the attempt to use online resources in the research environment was made, particularly including social networks, forums and online platforms that enable organizing information about actors, compiling the most interesting questions to ask them. One of the innovators in this area was L. Manovich who identified three key "data classes" when conducting research.

## METHODOLOGY

Their components are systematized in Table 1:

<b>Participant groups</b>	<b>Functions</b>	<b>Examples</b>
Data creators	Directly generate digital information	Internet users of a specific platform or social network
Actors - "digital asset" holders	Distribute data arrays collected on the basis of their own interests	Administrators of social groups, platforms, and websites under study
Actors having access to digital arrays	Use the Big Data predictive potential, and form projects and models of propaganda impacting participants of the first "data class"	Administrators of social networks, Internet websites and large Internet resources

At the same time, it should be understood that such a class division is based on topic modeling, which is only the outset of sociology or sociological research. Further, more in-depth focus should be made regarding the issues that the audience is interested to study. To form questions as accurate and significant as possible among the group studied, artificial intelligence came in its active use as early as in 2020, which enables assessing huge volumes of information by identifying the needs and resources of the audience under analysis. (Odintsov, 2017)

Thus, integration into Big Data sociology using artificial intelligence is a driver for modern sociological research method development, which is logically to consider based on a certain area.

The employment or labor market area is of interest, which is actively studied by sociologists and has high volatility depending on the influence of a number of external factors on it. Sociological research in this context is aimed at identifying the trend in profession priority ranking, change in the paradigm of applicants regarding the increased trends towards raising requirements to employers in the field of quick thinking, stress resistance, performance, teamwork skills and others.

Under these conditions, the need for new technologies was determined, which make it possible to determine market trends more accurately and at simultaneously build competence ratings mostly demanded on the market influenced by time parameters. Big Data is used in the research based on classification and regression.

Essentially, the task of classification is to identify the value of one of the of analyzed object parameters based on the values of other parameters. The definable parameter is a dependent variable and the parameters used to define it are independent.

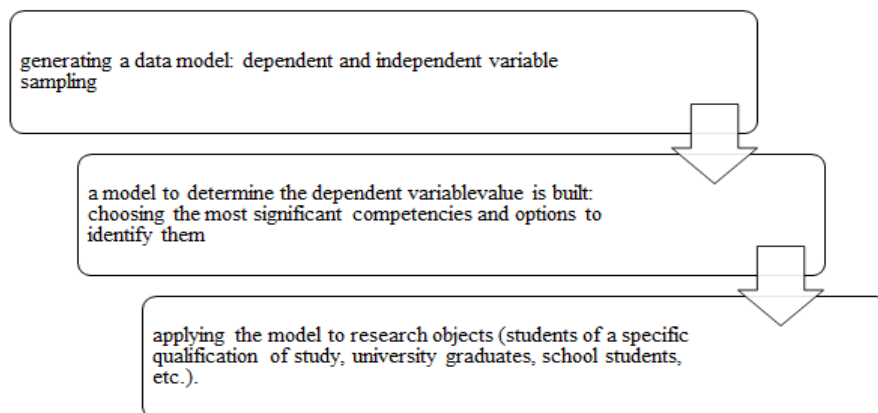
Thus, the labor market may have the following independent variables when carrying out research using Big Data technologies:

- Survey group age
- Family members
- Education
- Sector of employment
- Dependent variables are
- Specific competencies
- IT knowledge

In this case, the independent variable takes a finite collection value. (Zikratov, 2014).

## RESULTS AND DISCUSSION

The task to classify and regress groups that are labor market participants is solved in three stages, which are clearly reflected in Figure 1.



**FIGURE 1**  
**CLASSIFICATION AND REGRESSION OF GROUPS PARTICIPATING IN THE LABOR MARKET**

To build the most reliable research model, a number of requirements are imposed on this sample, the main of which should be systematized into a tabular form (Table 2):

<b>Requirement</b>	<b>Comments</b>
Large number of objects	More objects improve model accuracy
Sampling must include all groups represented in the segment	The maximum number of regression objects is covered
Sampling includes sufficient number of objects	Classification is done at the stage of each interval

When conducting a sociological research of the labor market, an important analysis area is to collect data from topic websites, Internet networks, communities which participants are potential objects to research. Such a tool is commonly referred to as Web-scraping. It enables the deepest analysis of huge amounts of text information, ensuring the simultaneous data flow from the studied web resources. (Khashkovsky, 2014.)

Big Data also uses search metasystems that consistently compare and select the results obtained from different sources (for example, the number of vacancies for a driver's position in Moscow, the most frequently mentioned competencies required for the vacancies under study, etc.). This method provides an opportunity to accumulate information when assessing labor market conditions by continuous scanning of news on market need changes (Zhuravleva, 2015).

The use of AI and big data in labor market research is a trend towards finding new opportunities to develop the labor market and new professions. In any case, the use of AI will optimize the research processes, while determining the market needs as accurately as possible.

Studying the experience of using Big Data in the framework of labor market research, the extensive research held by the consulting company Antal should be referred to. More than 6 thousand top managers and middle managers were involved in the survey. Most respondents (72%) work in Moscow and the Ministry of Defense. The smaller part (10%) work in St. Petersburg. In the survey, attention was paid to the change in competencies and the change in demand growth trend common to certain professions.

At the same time, Big Data technologies were used, which made it possible to build well-defined regression, highlighting the most relevant for employers staffing needs. In addition, a tendency was revealed indicating that instead of a profession, a modern person has a set of professional competencies and skills, which can be divided into groups given in Table 3.

<b>Skill group</b>	<b>Essentials of human skills</b>	<b>Examples</b>
Hard skills	Can be taught and measured	Typing on a computer, driving a car, reading, math, knowledge of a foreign language, using computer programs
Soft skills	Mental and interpersonal competencies	Sociability, teamwork, creativity, punctuality, mental ballast

In addition, in the course of the study, a trend was determined in changing the demand for professions among employers, leaving the market for some of them and development of new ones related to economic robotization, digitalization and digitisation. Under current conditions,

professions-algorithms develop gradually avoiding the same operation types, transforming them into human-controlled automated complexes.

Many new professions have also developed, that require new professional competencies corresponding to the labor market to be defined.

Based on the researches held to determine the change in labor sector thinking, the qualitative research also enabled defining an innovative trend in developing professional competences with employees. In the current context, professional competences of the applicant capable to effectively perform the functions are not always sufficient. Quite often, personal qualities dominate the ability to perform any given works.

The study made it possible to conclude that technical competence in modern conditions is ranked lower than first regarding the factors that affect employees' performance and employers' assessment. Interpersonal skills are key; their components are given in Table 4.

<b>Staff competencies</b>	<b>Description</b>
Judgment and decision-making	Judgment and Decision-Making
Service orientation	Advanced Emotional Intelligence
Negotiation	Negotiation Slikks
Cognitive flexibility	The ability to quickly switch from one thought to another, think about several things simultaneously, keep in mind heterogeneous and even contradictory ideas

Table 4 shows that several skills relate to the area of relationships. A number of competencies depend on the ability to think quickly, look at the essence, find a solution to problems, and generate new ideas. Under these conditions, people and their needs are to be studied more deeply in order to participate in the labour market, which is hardly ever possible using traditional techniques only.

However, it is worth noting that these studies are associated with a number of issues, including the inability to control reliability of information obtained from the survey participants, as well as inability to verify whether the participants really existed. Thus, when sampling randomly while using artificial intelligence, it surfs the Internet resources, reaching out to community and social network members that are somehow related to the study. At the same time, in some cases, it can trigger intentional distortion of facts by respondents who are involved in illegal activities and operate using fake accounts, which will reduce result reliability. To eliminate such a deficiency, samples should be monitored with sociologist's personal contribution, and applying selective testing of the results obtained.

In general, the advantage of data processing machine methods that use artificial intelligence is the possibility to study phenomena from the perspective of an array of qualitative characteristics. The obtained data enable identifying trends that are not visible in the system of standard sociological studies.

## CONCLUSION

The analysis draws to a conclusion that Big Data technologies are intensively used in sociology. Their key role is to process large data volumes with the resulted high validity in the short term. However, the greatest advantage of Big Data is that it enables ranking respondents by quality characteristics, revealing market trends. Regarding the labor market, these are criteria for

professional competencies, the requirements for which are continuously changing in the modern world, while the traditional parameters of market assessment are wages, age and status of respondents. This proves the importance of new technologies and increase in their capacity to carry out sociological research.

## REFERENCES

- Aharonovich, A.R. (2019). Socio-economic importance of state support for youth innovative entrepreneurship in the economic development of the state. *Academy of Entrepreneurship Journal*, 25, 1-6.
- Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 663.
- De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. In AIP conference proceedings (Vol. 1644, No. 1, pp. 97-104). American Institute of Physics.
- Khashkovsky, V.V., & Shkurko, A.N. (2014). Modern approaches to arrange systems for processing large data volumes. *Izvestia SFU. Technical sciences*. 8(157). 240.
- Mayer-Schonberger, V., & Kukier K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.
- Manovich, L. (2015). The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics [Electronic resource]. Retrieved from: [http://mano-vich.net/content/04-projects/088-cultural-analytics-social-computing/cultural\\_analytics\\_article\\_final.pdf](http://mano-vich.net/content/04-projects/088-cultural-analytics-social-computing/cultural_analytics_article_final.pdf) (access date: 23.01.2021)
- Odintsov, A.V. (2017). Public opinion sociology and the Big Data challenge //Monitoring public opinion: Economic and social changes. 30-43.
- Rodnyansky, D., Abramov, R., Valeeva, G., Makarov, I. (2020). Trends of Development of the Industry of Tourism after End of Coronavirus. *Revista TURISMO: Estudos e Práticas*, (4).
- Rodnyansky, D.V., Abramov, R.A., Repin, M.L., & Nekrasova, E.A. (2019). Estimation of innovative clusters efficiency based on information management and basic models of data envelopment analysis. *International Journal of Supply Chain Management*, 8(5), 929-936.
- Toshchenko, J.T. (2013). New trends in Russian sociology development. *Humanist of the South of Russia*. (1). 85.
- Zhuravleva, E.Yu. (2015). Modern sociology in a networked digital environment: from computational and electronic social sciences to digital social research. 5-34.