Social Aspects of Big Data Technology Implementation

Artem A. Balyakin[1], Sergev B. Taranenko[1], Marina V. Nurbina[1], Mikhail A. Titov[2]

[1] National Research Centre Kurchatov Institute, Moscow, Russia, 143968
[2] Lomonosov Moscow State University, Moscow, Russia,119991

https://doi.org/10.33847/2686-8296.1.1_2

That which has been, is that which is to be,
and that which has been done,
is that which will be done,
and there is no new thing under the sun.
Ecclesiastes 1:9

Abstract. Big Data is supposed to be one of the main traits of new coming digital era. Its technological aspects are usually widely discussed, whereas social peculiarities are mostly neglected. We present main approaches to Big Data, and argue that despite seeming revolutionary technology, Big Data can be treated as a new tool to produce knowledge. That means, it generates the same risks and challenges as other breakthroughs we witnessed previously. To our viewpoint, cultural aspects should be as counted as a main issue in Big Data implementation. Since the inability to control big data through prohibiting some peculiar features it possesses, we argue that one should focus on such practical steps as terminology improvements, and the evaluation of societal outcomes of the new technology.

Keywords: Big Data, Cultural Aspects, Scientific Infrastructure, Megascience, Research and Innovation Policy, Socio-Economic Challenges.

1 Introduction

With the pace of nowadays technology development digital aspects of everyday life become inevitable, and such a term is widely used to describe forthcoming a global information infrastructure. Experts predict the rise of "Internet of Things" that will rise from the current "big data", artificial intelligence and other similar technologies. As a whole, the most important thing is to put existing data in prescribed order, and in case it is duly collected, structured, and processed, the information would enable the society to arrive at right managerial decisions [1,2]. For instance, in the EU, the first task is to build a digital infrastructure that would ensure effective interaction between researchers and infrastructure elements [3]. In the Russian Federation, digital technologies are included in the number of breakthrough technologies, and their practical use in the future should contribute to Russia's global technological competitiveness [4].

Newly forming digital industry will be highly likely personal: it is expected to be characterized by a flexible network approach to the production process (so called network-centric approach), when each consumer turns to be a manufacturer, constructing the necessary goods upon one's purpose. In this scenario the industry of the future will be operating online, from sample designing to its production (through data exchange and numerical processing). There would be the geographical dispersion of various elements of production (i.e., production of the constituent parts of the goods). In optimistic conception, the innovation cycle would be also reduced. All these ideas are still under question, as the regulatory governmental function in the field of high technologies to increase.

With the development of electronic computer technologies and the Internet, the role of the management of large amounts of data is becoming increasingly prominent. The issues of transferring, processing and storing information from purely practical tasks of building hardware and software are transformed into a problem of infrastructure organization, and data handling issue is moving from technological solutions into the field of economics, sociology and public administration. Moreover, the evolution of "big data" technologies gives an impetus to the development of a number of specialized scientific areas, including dual purpose ones. In particular, we are talking about the creation of a system of highly specialized artificial intelligence (i.e., intelligent big data processing systems), the development of mechanisms for optimizing data selection using a statistical-probabilistic approach, and the creation of new methods for in-depth analysis of large amounts of data, methods for solving multidimensional incorrect problems.

The most common approach is to consider the Internet as a source of "big data", and, more narrowly, - to regard social networks as the only big data origin [5]. However, other areas of human activities, such as science, retail, and medicine play an important role in "big data" formation [1,6-8]. The apparent diversity of "big data" sources is being neglected by similar approaches in collection, storage and analysis of the gathered information. Hereby, the algorithms used in the scientific field can be transferred to other areas of knowledge.

Main feature of digital technologies development (and related areas) is a serious heterogeneity in its introduction and implementation into everyday life; at the same time, the dynamics of the information technologies dissemination in world regions does not meet optimistic expectations [9], the last ultimately leads to the emergence of new socio-economic and political challenges, the response to which requires increased attention of society, with the involvement of experts in various industry areas. Thus, one of the possible consequences is the information (digital) spatial inequality, which is irremovable in the short term [1].

In particular, one of the problems requiring a scientific approach is the legal implementation of processes of the "big data" circulation and the development of a common (more likely – cultural) approach to "big data" that takes into account the socio-economic and moral-ethical dimensions of new technologies.

In this paper, we consider the current situation of perception of "big data": in the first section a general description of the problem of "big data" is given, with corresponding definitions and concepts. The second section describes some features of the legal regulation of the "big data" circulation and existing practices, and shows the important role of the cultural aspect of the considered problem. In the third section, we postulate the idea of non-uniqueness of "big data", we provide parallels with already existing technological innovations in human history. Finally, in conclusion, a number of practical steps are proposed in the field of "big data" regulation.

Our main position is that the novelty of the practical use of "big data", as well as the problems produced by them, is very exaggerated: most of the difficulties have already arisen in human history throughout scientific and technological progress, thereby, we are able to assume (predict) the main consequences new technology can produce, and strive to take the most optimal management decisions in this area.

Based on the proposed approach, the conclusion contains brief summary and recommendations on the need and the possibility of "big data" regulating.

2 Conception of "Big Data"

The term "Big Data" does not have a conventional definition. Some researchers even exclude the technology of "big data" as an independent area, considering it as "…a title that includes a large number of technologies that are actively used in everyday life, related to the various areas of activity and do not have signs of innovation…" [1].

Previously, the main criterion for referring to "big data" was the amount of information processed, "the size of which exceeds the capabilities of typical databases for writing, storing, managing and analyzing information" [10], and the "big data" themselves were determined by specifying the following main characteristics of the operated data, usually referred as "Three Vs" (see Fig.1):

1) large volume (Volume),
2) diversity of data (Variety), and
3) high rate of their change (Velocity) [2].

In a broader sense, "big data" was understood as a socio-economic phenomenon associated with the emergence of technological capabilities to analyze huge amounts of data, in some problem areas - the entire volume of worldwide data, - and with the resulting transformational consequences [11].

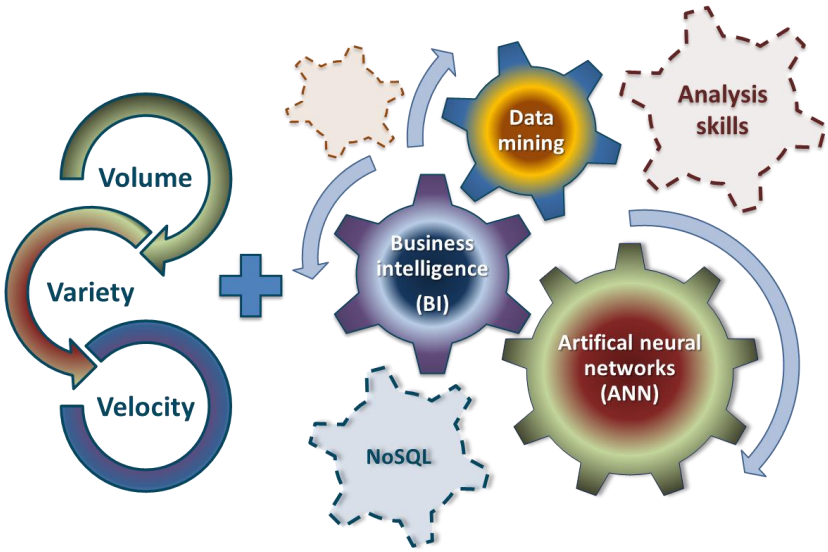

Fig.1. Big Data and Technologies

Gradually, the limitation of computing capacity for information processing began to wash out from this definition, and the main emphasis was placed on the methods and approaches to the processing of initial/raw data. By now, the main software methods used for the processing of "big data" are the means of mass-parallel processing of vaguely structured data, first of all, database management systems of the NoSQL category, MapReduce algorithms and software frameworks with corresponding libraries implementing the Hadoop project [12]. A variety of information technology solutions further began to be attributed to a series of big data technologies, to a greater or lesser extent, providing similar characteristics to the possibility of processing extra-large amounts of data. Fig. 1 illustrates the key technologies involved in "big data" operation, as well as main characteristics of "big data" as a process (Three Vs).

Since "big data" involves a combination of information, methods of its processing and obtained results (of the erroneously counted as new "knowledge"), the question of the source of data arose. Onwards we will utilize the term "data lake", that refers both to all the initial unstructured information in general, and the infrastructure that makes it possible to operate the "big data".

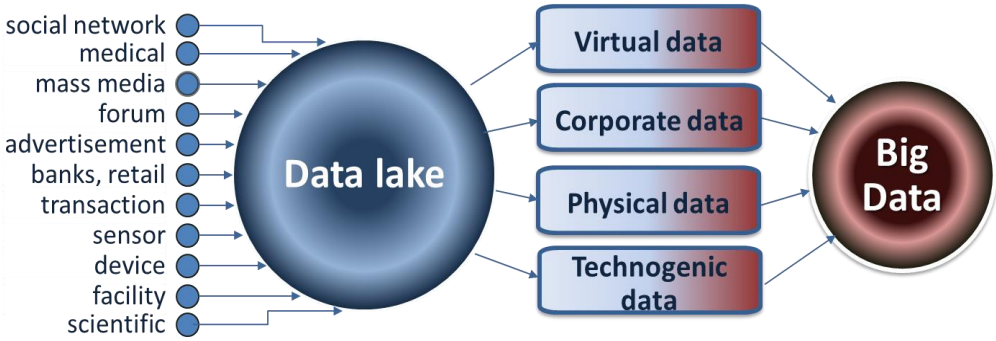

Fig. 2. Main approaches to Big Data. The current state-of-the-art

Since in the practical aspect the commercial use of "big data" is considered to be the most significant, we propose following classification according to the nature of original data (where from "data lakes" are formed/filled), represented in fig. 2:

- virtual data (e.g., Internet: social networks, forums, electronic media, etc.);
- corporate data (e.g., banks, advertisement, retailers - it is, usually, customer data and currency transactions);
- physical data (e.g., data from sensors, detectors, measuring tools and other devices);
- technogenic data (information exchanged between devices during operation; auxiliary information).

As it is seen, the resulting data set relates to various areas of human society: social (e.g., electronic data), economic (e.g., data of corporations and banks), medical (e.g., collection of personal data and a number of corporate data), scientific (e.g., physical data from scientific facilities) - which, in turn, leads to different prioritization in data handling and processing. Thus, for electronic data, the size (volume) aspect is important, for corporate data - the emphasis is put on variability (dynamics of data), while the data itself is often initially structured. For medical data, such parameters as volume and dynamism are important, but there is also a clearly postulated parameter "value of human life" (i.e. social and cultural aspect).

Practically, manipulation with "big data" requires the structuring information into categories (tags), and the corresponding accompanying information (answering to questions what, where, how, by whom, etc.) is often comparable in volume to the original one. Consequently, the rapid growth of the data archives leads to problems with scalability, efficiency of the monitoring system and delays in receiving a response to external requests. At the physical level, this results in the fact that working with "big data" means [partial] segmentation of data on tasks (in accordance with the metadata model adapted to the requirements of the system: monitoring of terrorist activity, updates in the Facebook, analysis of scientific data, etc.)[1].

The requirement for computational efficiency (e.g., minimizing the amount of data processed in memory while generating a report) leads to the need to generate reports at different levels of detail [8]: in fact, "big data" is a data model with stepwise aggregation (in the limit – infinite recursion, and thus, fractal).

In the future, we will consider "big data" as a kind of entity with the following properties and meeting the following criteria[2]:

[1] To obtain more advanced position, investors/shareholders may have the interest to assess more detailed and structured information extracted from "data lakes". This is especially crucial for technology-oriented industries supposed to be capable to convert R&D expenditures into innovative products. This can be a key driver of competitive advantage and hence the financial performance of the organization. For instance, meaningful patent indicators can be an interesting proxy for assessing a firm's capability to innovate and to gain competitive advantage. Lexis Nexis can be cited as an example of a company that evaluates the competitive advantages and vulnerabilities of firms using a big data analytical platform [13].

[2] As an example of "big data" handling a one can refer to "data lakes" operating chart at Tinkoff Bank (see for details [14])

- The initial data set is considered to be some "data lake": unstructured, dynamically changing, heterogeneous and loosely coupled.
- "Big data" is the result of applying some methods (models) of selecting data from the all information available in the "data lake", and includes both the data itself with corresponding categories, and the mechanism (algorithm) of their selection and analysis.
- "Big data" is a dynamic, non-stationary system that is in constant process of filling, updating and adjustment[3].
- Inside the "big data" is embedded the algorithm of their processing, issuing a response upon some request. In this case, a feedback mechanism is implemented[4]: the data obtained, in turn, modify the "big data" and/or the mechanism of their selection from the data lake[5].
- Any "big data" is scientific in the sense that it is measurable, selected by the chosen model, processed in advance by given methods and algorithms.

Towards the creation of a digital infrastructure that uses large amounts of data, several problems are highlighted, both of technical and institutional nature, in particular [15]:

- Data encoding;
- Elimination of "garbage", data "noise" management (e.g., filtering out unnecessary information, extracting useful information, evaluation of data adequacy);
- Addressing issues of the long-term content preservation, the development of new storage devices, backup technology;
- Compatibility of data from different periods of time (methods of data writing and encoding are different today and 10 years ago). Compatibility of data from different fields of science;
- Numerous duplications and repeatability of data, information redundancy;
- The need for continuous data verification and its "repackaging" (saving in a more compact form and/or more accessible). Data reduction (for writing) and their recovery (for adequate performance on request);
- Data presentation (visualization);
- The problem related to the creation of metadata: the transition from simple records to complex ones, having external and internal references for navigation (e.g., the analogy to the Internet); metadata structure development;
- Organization of data search and retrieval: formalization of search queries, caching of search data, allocation of servers for storage depending on the tasks and stored information. Organization of multi-level data access (e.g., the analogy with the library);
- Compliance with the legal issues of storing information, which is related to different countries (jurisdictions);
- Resolving the issue of territorial distribution of stored information.

Thus, we can distinguish "big data" issues in following aspects:

- man-made (problems associated with the technical evolution of storage devices, communication channels, etc.);
- structural (associated with duplication, redundancy of information, the creation of metadata, and TP);
- organizational and legal.

In our opinion, the most important problem to date is not technical difficulties, but issues of regulating "big data" circulation, providing for both following national interests and ensuring the protection of human rights (in case of using and processing personal information). In that question, science provides examples of both the appearance of the

[3] Any arbitrarily large structured data is just a large database.
[4] Data should be continuously updated or modified (including its structure).
[5] In practice, it can be said that the data at the time of its retrieval from the "data lake" is ALREADY outdated.

first difficulties of the legal regulation of "big data" and the emergence of moral and ethical problems, as well as possible ways of addressing them.

3 Cultural Approach to "Big Data"

The fact is that modern science and technology are inseparable from the socio-economic and political life: all areas of activity are so intertwined that it is impossible to separate technological progress from changes in social norms. In the formal language of title deeds, this led to the fact that, for example, in the European Union, the basis of decisions made is the need to solve social and humanitarian challenges in all their manifestations [3,15]. With respect to the "big data" in the EU there is a discussion about the ways of their management and regulation; Biomedicine (Artificial Intelligence for Decision Making) was selected as the first field of application: metadata is now being collected and ethical principles relating to the regulation are being developed.

In the United States, the first concerns are not about controlling the circulation of "big data", but about the technical access control: the so-called "neutral" Internet involves changing the physical parameters of access to information and its processing in accordance with its content; in practice, data and users of these data are ranked due to their status. In the United States, issues with the protection of private information and its transfer to third parties for analysis and processing also led to a number of serious scandals with social network companies, but it was not possible to formulate clear definitions of permissible information disclosure (a discussion of existing legislative initiatives is given in [16]).

In Russia, the topic of "big data" has not gone beyond the highly specialized approach yet: thus, it raises the question of the need to create a public operator, allowing private sector participation, who would manage information about users' social data (e.g., user preferences on the Internet, social connections, the circle of communication, etc.) [17]. At the same time, legal issues are mainly limited to the regulation of access to personal user data by third parties [5,18] or the use of collected information to solve legal problems (e.g., user localization on the basis of geo-tracking of his mobile phone) [19]. Evidently, "big data" are a key interest for financial authorities, who are eagerly striving to get access to private users' information in order to streamline taxation[6].

The first problem that arises in the field of "big data" circulation is connected with personal data protection. Main issue is the necessity of personalized consent (confidence agreement) that includes the information to which extent the consent is given, and the procedure for its use. Practically, this leads to the emergence of a new segment of "big data" (the volume of regulatory legal documents, with the details of the base to be regulated, is comparable with the initial volume, i.e., it is also "big data" in volume). On the other hand, the set of all available data is so large (the size of data lakes is significant) that even without their personalization it becomes possible to identify the individual without any doubt, i.e., "data protection" function becomes useless. A way to break this vicious circle could be classifying "data lakes" as "natural" (i.e., data with no possessor, belonging to anyone), as suggested by the McKinsey report [21].

It is also important to note that the existing legal restrictions on the processing of personal data solely in accordance with the originally stated processing objectives, as well as the inadmissibility of combining different databases with originally stated and incompatible processing objectives, contradict with the existing technology and business practices, since it eliminates the advantages provided by "big data" technologies [22].

Note that a superficial way on the issue of regulating "big data" circulation leads to the thesis about the prohibition of social networks or their artificial restriction by introducing forbidden words, topics, etc. However, this approach is doomed to failure according to the above concept of "data lakes" and "big data": the initial data per se can be any, and their selection and analysis play the most important role (nor its origin or content). As practice shows, while maintaining and enhancing the existing attitude to the information regulation, "big data" technologies will increasingly migrate towards

[6] The last attempt of "big data" regulation was undertaken in newly proposed draft law on the creation of a single information resource for all citizens of the Russian Federation and persons residing in the territory of the Russian Federation [20].

DarkNet, thus, exiting from the legal field. Thereafter, it should be about regulation at a level higher than the formal prohibition of certain words and/or pictures.

For example, in the EU, the concerns expressed about the recently expanding requirements for the protection of personal data will lead to the suspension of work in the field of "big data". As an example, the General Data Protection Regulation (EU) is provided (2016/679, EU GDPR) [23]. This block of laws is aimed at giving citizens the control over their own personal data, at the same time suggesting the simplification of the regulatory framework for international economic relations by unifying regulation within the EU. The law expands the concept of personal data, introduces the concepts of "cross-border data transfer", "pseudo-anonymization", establishes the "right to oblivion", introduces the role of a security officer. The continuation of this policy is the recently adopted directive on the protection of copyright (EU Copyright Directive).

Another problem is the task of storing and accessing "big data". In general, a more capacious data warehouse, an improved search system, maximum complementarity and coherence of information are offered. In order to obtain maximum results and implement the right of equal access, the EU actively implements the principles of open science: thus, 779 organizational declarations governing open access were noted in the ROARMAP report for 2016. Of these, 133 were prepared by investors, 636 were formulated by scientific organizations. The "open science" movement in the Russian Federation has not become one of critical technology yet, but the creation and promotion of "open" principles for building digital infrastructure is included in the number of recommendations for adjustments to the scientific and technological priorities of research and development. Thus, in the foreseeable future, it is planned to implement a system of distributed remote access to both unique scientific facilities and databases, the development of systems for cloud computing and information storage.

This, in turn, leads to the fact that "big data" creates the illusion of knowledge, when the quantity replaces the quality. Easily accessible information leads, for example, to the desire to launch an "unconditional digital income", which contradicts to the fact that most of the information is useless and is never used (e.g., the CISCO report indicates that the "data lakes" themselves are growing at an enormous rate and according to expert estimates, at present, up to 90% of lakes are useless, since they are overfilled with information collected for some unknown purposes [1]).

4 Big Data: Totally New Deal?

The list of issues stated above are dangerous, challenging, but in fact nothing new for the society has come into life: such concerns, as can be easily seen, have already arisen in human history. The key point can always be reduced to the inquiry, if new technology introduces new knowledge, or just provide the new solution (assess)? If "big data" challenge is unique? Our answer is "sure not". We recall, for example, the invention of typography and the change in the status of monasteries at the end of the Middle Ages, when they lost their functions of knowledge accumulation and preservation.

The first issue is connected with the volume of information. It puts the question of data preservation and transference, rather than the extraction of new knowledge. In fact, we are witnessing the return to the present of the difficulties and problems of the middle ages: how to properly organize the functioning of libraries? Storage and preservation of manuscripts? How we can sure that their rewriting by literate monks?

The amount of data (information) can increase, but quantity doesn't turn into quality. For "big data" it is widely accepted the concept of easy search for correlations: we find what happens after what, while not explaining, why. We also note that data/information (not knowledge!) has become more accessible, their spread has accelerated significantly. But whether it is really new knowledge? The amount of information created today generates the same scholastic problems as before: does new knowledge exist or does the overall objective only narrow down to the correct codification of the existing one? To date, the answer is most likely negative: quantity (amount of data) does not yet turn into quality. However, there is a tendency of the dilution of the notion of "big data", when society's demand shifts from quantity (size) to quality (algorithms and results of the technology use).

An obstacle to the development of the technology of "big data" is, in our opinion, a bias against it: the strong term is shifted from the field of scientific analysis to the field of hype. The main problem now in the field of using "big data" is the lack of a culture of handling a new tool of scientific and technological progress. Therefore, many problems and difficulties arising in the application of the "big data" technology do not have a conscious cause and malice: they result from ignorance (both the mechanisms and algorithms of the "big data" and the interaction of high technologies with the society). Another problem is associated with seemingly redundancy, accessibility and "belonging to no one" knowledge.

The lack of a culture of handling "big data" means no practices in the relevant field. Accordingly, at this stage there is no object for legislative work. In this regard, the expert community can and should form these practices.

In this case, the holistic approach is important. Currently, the most common attitude involves answers to ongoing challenges, the regulation of particulars: restrictions and prohibitions are created, which in fact are very easy to get around - this is not about lacunae, but about holes. The whole experience of mankind shows that such an approach is doomed [to failure]: an attempt to manage "big data" "piece by piece" creates new "big data" that governs the original "big data": it turns out to be a vicious circle.

We argue that it is required to perceive "big data" as a new tool of conceiving the world, carrying both positive and negative sides. This leads us to the conclusion that it is necessary to manage the social dimension of high technology. "Big data" technology accompanied with "open science" approach create endless "data lake" set. It brings us to the need to find a new way to data flows regulating. We propose to focus on "props of meaning" (main concepts, ideas, algorithms, physical basis of "big data" technology such as Data-centre), rather than on virtual space with unformed social communication practices.

In addition to this "from above" approach, the induction method is also possible. For example, one of the options for resolving such problems is the case law, which fixes the established tradition. Practically, society is waiting for some event to happen (e.g., the Cambridge Analytica data scandal) in order to begin to regulate this field. Until a certain moment, there is a fear of the new and unknown, to which unique (dangerous) features are attributed. It does not take into account the fact that similar problems have already arisen earlier, and it is just required to adapt them, "translating" from the old language and terminology into a contemporary perspective. As an example, we recall the dispute about the responsibility of artificial intelligence used in self-driving cars. It is proposed, for example, to use the Roman approach and to consider AI as an analogy for a slave (servile) in Roman law, i.e., not a subject, but an object of law. In such interpretation, there is no need to expand the concepts and introduction, as some researchers suggest, of a new - digital personality (which AI would be endowed with).

As the authors see it, the development of "big data" technology still raises more questions than answers, which is why it is now important to promote the development of new technologies with taking into account their social consequences. It is, in our opinion, the equivalent to the formation of the scientific culture of using digital infrastructure. Those are people who write the rules and set the language of the future. Its symbols are the digital infrastructure, but the logic of the organization of communication will be set by human. This is a hermeneutic approach: a well-formed language structure solves half of the problems.

This idea can be regarded as a possible treatment of approaching Digital Age, as well as a format of social science in the Digital Age. We suggest it to be a kind of a New Deal with regards to "big data" technology and its application.

5 Conclusion

In our opinion, the problem of "big data" represents an important task of modern social science: the versatility of methods and approaches in the technology of "big data", the "trans-border nature" of their consequences, the level of impact on the society - all this makes us consider new technologies both as an opportunity and as a serious challenge at the same time.

Big Data technology is an inevitable tool for digital economy that is being formed now. Its implementation requires the solution of both technical and social tasks: moreover, socio-economic impact of big data is often neglected, and needs to be addressed. With this regard it is extremely important to apply to previously arisen problems, adapting former answers for modern risks and challenges.

By now we witness the serious overestimation of the "big data" technology, when it appears to be something revolutionary that drastically changes human nature. In our previously published work [24] we stressed that the novelty of big data technology is mostly semblant, and the human society has already undergone similar problems earlier. We argue that this misconception is rooted in "trendy" perception of digital science as a whole and big data particularly. In fact, much of the moral and ethical issues have already arisen to humanity in the past, and the most important way to solve them was not partial prohibitions, but an understanding of the problem as a whole, developing a culture of dealing with a new phenomenon.

The authorities have 3 ways to influence the development of big data technology: to regulate it (via standards, etc.), to prohibit it (the most adequate example – GDPR introduced in EU [23]), and to stream it via infrastructure development (i.e. constructing data centers, etc.).

The role of experts' community is to raise questions of probable socio-economic issues in advance, to find the very position of big data technology in a whole structure of digital society. As for the practical steps, authors would suggest:

First, the coordination and promotion of a single glossary in the field of "big data", taking into account the experience from various fields of activity.

Second, the rejection of attempts to control the information used by "big data", and move the emphasis on the result of their use. It means the shift of the focus from "data lakes" to other objects (the conclusions drawn from "big data" technology, the regulation of data centers, etc.).

Third, bridging the gap between the technologies themselves and the consequences of their use. This requires taking into account social and economic effects of high technologies.

Neglecting social outcomes of new digital approaches could be the main problem in Big Data technology implementation. Reducing all problems solely to technical issues should be avoided, and emerging digital society should take complex form of advantageous combination of high-tech solutions and high-tech culture (i.e. culture of handling new technological solutions based on previous experience).

Acknowledgments

This work was supported by RFBR grant № 18-29-16130 MK.

References

1. Hype Cycle for Emerging Technologies (2015), URL: https://www.gartner.com/doc/3100227/, last accessed 2019/03/31
2. Lynch, C.: How do your data grow? Nature 455, 28-29 (2008).
3. Florio, M., Sirtori, E.: Social benefits and costs of large scale research infrastructures. Technological Forecasting and Social Change 112, 65-78 (2016).
4. The decree of the President of the Russian Federation. About the strategy for the development of the information society in the Russian Federation for 2017-2030. No. 203, 09.05.2017
5. Almyrzaeva, A., Kostyuk, V., Nevredinov, A.: The role of Big Data in modern society. Journal of Economy and entrepreneurship 9 (3), 580-582 (2017).
6. Yuchinson, K.: Big Data and Legislation on Competition. Law. Journal of the Higher School of Economics, issue 1, 216-245 (2017).
7. Grigorieva, M., Golosova, M., Ryabinkin, E., Klimentov, A.: Exascale Store for Scientific Data. Open Systems. DBMS. 23 (4), 14-17 (2015).
8. The Global Information Technology Report. The World Economic Forum (2016), URL: http://www3.weforum.org/docs/GITR2016/WEF_GITR_Full_Report.pdf, last accessed 2019/03/31
9. Zobova, L., Shcherbakova, L., Evdokimova, E.: Digital spatial competition in the global information space. Fundamental research, issue 5, 64-68 (2018).
10. Manuka, J., Chui, M., Brown, B., Bughin, J., Hobbs, R., Roxburgh, C., Byers, A.: Big data: The next frontier for innovation, competition, and productivity / McKinsey Global Institute (2011).

11. Mayer-Schönberger, V., Cukier, K.: Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt (2013).
12. Making Sense of Big Data. PwC Tech Forecast, issue 3 (2010).
13. H. Ernst, N. Omland The Patent Asset Index – A new approach to benchmark patent portfolios: World Patent Information 33 (2011) 34–41.
14. https://habr.com/ru/company/tinkoff/blog/259173/
15. Balyakin, A., Mun, D.: Formation of an open science system in the European Union. Information and Innovations. Proceedings of the conference "Scientometrics and Bibliometrics", pp.33-37 (2017).
16. Net Neutrality 2019 Legislation. National Conference of State Legislatures (2019), URL: http://www.ncsl.org/research/telecommunications-and-information-technology/net-neutrality-2019-legislation.aspx, last accessed 2019/03/31
17. Hu, J., Zhang, Y.: Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization. Scientometrics 112(1), 91-109 (2017).
18. Sokolova A.: The impact of big data analysis technologies (big data) on personal data legislation. In the collection : Jurisprudence 2.0: a new look at the right materials of the interuniversity scientific-practical conference with international participation. Russian University of Peoples' Friendship . Moscow , 2017. pp. 282 -285.
19. Bulgakova E.V. Methods for analyzing big data in solving legal problems. In the collection : Law and Information: Questions of Theory and Practice, a collection of materials of the international scientific-practical conference. Ser. "Electronic legislation" of the FSBI Presidential Library named after B. N. Yeltsin. 2017. pp. 90 -96.
20. Order of the Government of the Russian Federation of July 22, 2019 No 1627-p On submitting to the State Duma of the Federal Assembly of the Russian Federation a draft federal law "On a single federal information resource containing information on the population of the Russian Federation", http://government.ru/dep_news/37452/, last accessed 2019/07/30.
21. Digital Russia: a new reality (2017), http://www.tadviser.ru/images/c/c2/Digital-Russia-report.pdf, last accessed 2019/03/31.
22. Savelyev, A.: The Issues of Implementing Legislation on Personal Data in the Era of Big Data. Law. Journal of the Higher School of Economics, issue 1, 43-66 (2015).
23. The EU General Data Protection Regulation (GDPR), https://gdpr-info.eu, last accessed 2019/03/31.
24. Balyakin A.A., Malyshev A.S., Nurbina M.V., Titov M.A. (2020) Big Data: Nil Novo Sub Luna. In: Antipova T. (eds) Integrated Science in Digital Age. ICIS 2019. Lecture Notes in Networks and Systems, vol 78, pp. 364-373. Springer, Cham.