BIG DATA AND ALGORITHMS IN THE PUBLIC SECTOR AND THEIR IMPACT ON THE TRANSPARENCY OF DECISION-MAKING[1]

Gergely László Szőke[2]

DOI: 10.24989/ocg.v331.25

Abstract

Big Data is clearly one of the most used buzzwords nowadays, but it really seems that the phenomenon of Big Data will have a huge effect on many different fields, and may be regarded as the new wave of the information revolution started in the 60s of the last century. The potential of exploiting Big Data promises significant benefits (and also new challenges) both in the private and the public sector – this essay will focus on this latter.

After a short introduction about Big Data, this paper will first sum up the potential use of Big Data analytics in the public sector. Then I will focus on a specific issue within this scope, namely, how the use of Big Data and algorithm-based decision-making may affect transparency and access to these data. I will focus on the question why the transparency of the algorithms is raised at all, and what the current legal framework for the potential accessibility to them is.

1. Big Data – the new wave of information revolution

The expression of "Big Data" is definitely one of the most used buzzwords in any discussion about recent technological development. Despite or besides the "Big Hype" about Big Data[3] it really seems that this revolution "will transform how we live, work, and think".[4] The recent tendencies can be regarded as the new wave of the information revolution started in the 1960s.

The Big Data phenomenon has many faces, no comprehensive definition is used – the approach by IT experts and tech companies, by economists or by lawyers may be different. Still, there are some common cornerstones to describe the Big Data tendencies: the industry uses the so called 3-4-5 (or even more) "Vs", as decisive features:[5] (1) Volume refers to the enormous and fast increasing

1 SUPPORTED BY THE ÚNKP-17-.4. NEW NATIONAL EXCELLENCE PROGRAM OF THE MINISTRY OF HUMAN CAPACITIES

2 Assistant professor and head of the Group of ICT Law, Department of Administrative Law at the Faculty of Law, University of Pécs, and researcher of the Big Data Research Group of the Szentágothai Research Center of the University of Pécs. E-mail: szoke.gergely@ajk.pte.hu

3 Cf. for instance Steve Dodson: Big Data, Big Hype? https://www.wired.com/insights/2014/04/big-data-big-hype/, but this question has been raised several times in the past years.

4 Cf. Kenneth Cukier and Viktor Mayer-Schönberger's famous book: „Big Data: A Revolution That Will Transform How We Live, Work, and Think" (2013, Houghton Mifflin Harcourt)

5 Cf. Gartner's definition containing volume, velocity and variety (https://www.gartner.com/it-glossary/big-data), the IBM's approach of 4V (http://www.ibmbigdatahub.com/infographic/four-vs-big-data, which was later completed with the fifth V of „Value" (https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/), and even complemented with Variability and Visualisation (by Eileen McNulty http://dataconomy.com/2014/05/seven-vs-big-data/)

amount of data; (2) Variety means the diversity of the data (including the fact that most of the data are unstructured, and could hardly be processed with the "old" data mining techniques); (3) Velocity refers to the fact that both the appearance of any new data and the expected analysis of these data are very fast (often real-time); and (4) Veracity points to the uncertainty of data.

Data alone are not enough, of course, so the main issue is how to analyse them. Not surprisingly new data mining technologies and powerful algorithms have emerged recently to solve this problem, including self-learning algorithms, which may "develop" by themselves, if enough data are available for them. But the operation of the algorithms is often described as a "black box". First, the actual way how they work is not understood by average people, and, besides this, the algorithms are usually treated as a treasured trade secret of the tech companies. Moreover, machine-learning mechanisms may pose a further risk, since their operation and their results are often inscrutable at times even for their programmers [5, p. 10].

Some general consequences of the spread of Big Data analytics are predictable. As Cukier and Mayer-Schönberger summarize it, first of all, no sampling of data will be necessary, since all (or almost all) data can be analysed – as they refer to it, using the terminology of statistics, "n = all". Once we have so many data, no clear hypotheses or straightforward questions are needed. This also means that unexpected and surprising correlations may be revealed. Another consequence – say Cukier and Mayer-Schönberger – may be the shift from causation to correlation: Big Data analysis only answers the question: "what", but says nothing about "why", but in many situations this is far enough [6, pp. 30-32]. Finally, and from a legal and ethical point of view this seems to be the most important one, Big Data analysis tends to be used for predicting future actions and behaviours with a considerable degree of probability (and predictions can be easily used for manipulation and nudging) [25, p. 25].

Generally, it seems that exploiting the potentials of Big Data holds out unprecedented business opportunities and benefits for the society as a whole, but also gives rise to new risks and challenges, of course. The potential spheres of usage are almost endless: development of new products and services, more effective (online) marketing and sales activity, health care, energy networks, transport and traffic systems, journalism, crime prevention and investigation, urban development, or gaining new scientific results, etc.

2. Big Data in the public sector

2.1. Potential use of Big Data and algorithms in the public sector

It is quite avowed that the activity of the public sector is based on data: using information is deeply embedded in the services provisioning, inspection and policy-making activity [14, p. 363]. So any changes (mainly if these changes seem to be radical) in the way of managing data shall significantly affect the operation of the public bodies.

As for the positive effects, Big Data analytics in the public sector may result in smarter data management, more effective (evidence-based) decision-making, personalized public services, better predictive analysis and problem solving [11, p. 386], and it may reduce fraud and corruption, increase transparency [20, p. 289], – so generally more productivity and efficiency is expected. These are more or less the same opportunities that come up in the business sphere.

Although there are some driver factors, there are also significant constraints of the spread of Big Data analytics in the public sector, like the lack of political willingness, lack of skilled people or legal uncertainties and concerns [18, p. 199]. Still, if we have a look at the relevant literature, we may find many examples and efforts for the usage of Big Data and algorithms in the public sector from all over the world: some of these are only plans or intentions, some others are in a pilot program phase and there are also examples for everyday use. Before looking through these examples and ideas, it has to be noticed that some of them have emerged in countries with a much lower level of privacy protection compared to the strict European data protection regime. However, I find it useful to show these tendencies too, to make the potentials (including the potential risks) of Big Data analytics more tangible.

1. Analysing traffic and transport data. Big Data analytics is often used for making traffic and transport systems more effective. In Brazil, for example, the continuous monitoring of traffic and road conditions helps to reduce the time to identify traffic problems from several hours to several minutes and assists in prioritizing road repairs. In Japan, an integrated system is designed to resolve traffic problems, like traffic congestion and accidents [11, p. 387]. If we use our fantasy, many other solutions can be imagined (some of these are likely to be used somewhere): analysing the data of all passengers and travels may help fine-tuning the mass-transit system (e.g. instant decisions on starting another bus or tram), analysing of real-time traffic data may help using intelligent traffic lights, or – by analysing the bigger picture – may help decide where a new road should be built or which older one should be reconstructed.

2. Improving health care services. Big Data holds out great opportunities both for more effective medicine and for improving the health care system itself. Analysing a huge amount of data may help detecting harmful (or even deadly) drug interactions [4], and may improve personalized medicine [1]. Predictive analysis, as it has been tried in Australia, may help in the hospitals' resource management, such as bed management, staff resourcing, and scheduling of elective surgery, but even the workload of an emergency department can be predicted with an accuracy of up to 93% [11, 13].

3. Fighting against tax fraud and corruption. We may find some examples that new data mining technologies are used in tax administration [22] and fighting against corruption, which seems to be reasonable, since a quite complex approach and analysis of a wide range of data is needed in these fields. There was, for instance, a Hungarian pilot program for using text-mining methods to analyse text-based public databases (procurement database, company registry, legal databases and publicly available other sources, like forums, blogs, and social media activity) in order to find fraudulent practices in the public procurement processes. The semi-automated system helped to find signs of suspicious behaviour, like invalid bids, co-ordinated high prices, geographically based market co-ordination, suspiciously similar prices or tenders, etc. [23].

4. Big Data for crime prevention and in criminal procedures. It is generally accepted that prevention is always better than reaction. Using Big Data methods makes predictive policing available, where historical (crime) data is used to discover trends and patterns, which might allow more effective and efficient deployment of police forces [18, p. 198]. Former data may be used to predict the location and time of a future crime likely to be committed, including identifying endangered zones [6, 19]. However, prediction may be used for a single person too. In some states of the United States software algorithms are used to assess some risks regarding the defendants, like pre-trial risks and risks of re-offending – and this information is used in the procedure or in the final judgement by the court [8, 15]. We could also find examples for actions taken against people before

committing a crime. In 2014, the "Chicago Police Department sent uniformed officers to make »custom notification« visits to individuals," because they had been identified by a software as people likely to commit a crime. "The idea was to prevent crime by providing them with information about job training programs, or let them know about increased penalties for people with certain backgrounds" [17]. Besides these examples of predictive analysis, it is important to see that Big Data methods are also used in the course of the investigation, e.g. for mapping the complex connection graph of the suspect [19, p. 192].

5. Fighting against unemployment. At first glance, it is not an obvious field of use, but we may find some examples for using Big Data and algorithms to reduce unemployment. In Germany, the historical customer data (including profiles) were analysed by the German Federal Labour Agency in order to offer more personalized services for unemployed people [11, p. 387, 18, p. 198].

6. Big Data in education. Among many other possible fields, algorithms may be used for complex evaluation of students' school performance [9, p. 1], but, on the other hand, also for analysing a mass of social media messages in order to get a real picture about the students' satisfaction with their courses and teachers [21, p. 17].

7. Analysing environmental data. Mass collection and analysis of environmental data may happen for several reasons. In India, for instance, real time monitoring of water flow is planned to minimize unaccounted water by detecting large changes in water flow [11, p. 387]. In the United States, algorithms help asses "the risks children face from exposure to lead at hazardous waste sites". A software for weather forecasts also uses a huge amount of data and sophisticated algorithms [9, 12].

These more or less randomly collected examples clearly show the relevance of Big Data analytics in the public sector in a very wide range and in divergent fields of use. It is clear that the public sector may win much by using the new technologies, but some of the examples might also remind us about the potential risks and challenges.

2.2. Risks and challenges

Using Big Data and algorithmic decisions may give rise to many ethical and legal concerns. If we think about taking any kind of actions against a person who is only predicted to commit a future crime or the fact that Big Data analysis may easily reveal hidden, surprising correlations and new information "never asked for", the alarm is surely sounding in many lawyers' heads.

The rich legal literature on these issues focuses – besides some others – mainly on discrimination, on privacy concerns, (both are very important and valid, but not the subject of this paper now), and on transparency and accountability for the algorithmic decision making. I am going to continue with this latter issue.

3. Big data and transparency of algorithmic decision-making

Although it seems at first glance that everyone may have easy access to information via the Internet and online services, access to the Big Data (meaning access to a huge amount of data), and especially, access to the new data analysing technologies and algorithms is actually limited. The present era "is characterized by an increasing concentration of the control over the information in the hands of a limited number of private and public entities, which, in different cases, cooperate in

sharing information with each other to increase their position as owners and gatekeepers of knowledge" [16, p. 23]. The imbalance of the informational power seems to be bigger than ever.

I collected many fields of use of Big Data analytics and algorithms in the public sector, providing some actual examples. It is quite clear that the results of any decision based on these new methods and techniques may significantly affect the life of a single person or a smaller or bigger community, but sometimes also the whole society.

Theoretically, it may have many advantages if the decision is based on a huge amount of real data and on smart algorithms. First, it seems that the exclusion of human nature may lead to a more objective and unbiased result. In many cases this is far from true, and the algorithms may contain biases. This may be "pre-existing", which means that if the used data show a discriminatory practice, the algorithm will "learn" this and use it as a standard. There may also be hidden problems in the code itself, which may lead to biases, and these may be invisible even for the programmers [9, p. 2].

Another problem that may arise is that the use of Big Data and algorithms may show surprising, unexpected results, usually only showing the correlation without justified causality. How can these kinds of "unjustified results" be accepted by the subject of the decision? [24, p. 108]. How can someone challenge a decision which is not based on reasons, only on correlations, especially if the way how the result came up is not known at all?

If the algorithms are not transparent, some further uneasy questions may arise: How is it known that the algorithm is fair and just, and not discriminatory? What are the limits, how is it known when it will break down or fail? What are those data that are excluded or overemphasised by the algorithms? [7, pp. 9-10].

Based on these reasons many argue that algorithms should be more transparent, or somehow controllable, or at least these issues have to be seriously dealt with [7, 9, 10, 16]. On the other hand, it also has to be noted that the development of new and powerful algorithms is the driving force of today's IT innovation, the success of companies nowadays, both that of the big ones and thousands of start-ups, is largely based on the new analysing methods. In my view, it is reasonable to distinguish between the decisions in the business sector and in the public sector. In the business sector the relationship between the partners is – at least theoretically – based on voluntarity, and ideally the market competition works well enough to force business actors to offer the most desirable terms and conditions, including as much transparency as expected by the customers. So generally the decisions of the companies, for example, a bank's decision about a loan based on a scoring system, or about the amount of the insurance fee based on certain personal circumstances [24, p. 108] are not subject to detailed justification by law. Admitting that this is an idealized picture also about the business sector,[6] in the public sector the situation is inherently different. The decisions are usually compulsory for those concerned, and there is no possibility to turn to the "competitor" if the "terms and conditions" are not acceptable. On the other hand, procedural rules are guaranteed by fundamental rights and by detailed legal regulation, and there is a possible remedy and judicial control over the decisions. However, to use any possible remedy, information is needed about the details of the background of the decision, so that is why generally it has to be

[6] Some companies are in a monopoly or quasi-monopoly position (e.g. due to the network effect), or all the actors on the given market basically use very similar data analytics while providing their services, etc.

justified in a written form by the decision making body. As a summary, the transparency of the algorithms seems to be much more important in the public sector than in the business sphere.

Finally, accepting at least the thesis that transparency of the algorithms has to be studied, and seeing that the legal situation concerning the transparency of algorithms is far not obvious, as a first step it is worth to give an overview of the affected legal institutions and instruments. Therefore, I will try to provide a picture of the relevant legal provisions.

4. Current legal framework of the transparency of algorithms

If we are thinking about the transparency and accessibility of certain information (the algorithms) in certain decision-making processes, it is first worth to study their accessibility by the general public, and then the possibility to have access for those who are the subject of the decision.

4.1. Access to algorithms by the general public

The first question worth studying is whether the algorithms on which a decision is based are accessible for the public under Freedom of Information (FOI) Law. The detailed rules of freedom of information and/or access to public documents are not harmonized in the EU, so they may vary in the different countries. So the starting point of the analysis is the Hungarian FOI regime,[7] but my consequences are valid in a much broader sense, since it seems that the basics of FOI Law, namely what is accessible (in most countries the data or information, no matter what medium it is stored on), and the limits of access (national security, personal privacy, commercial confidentiality and internal documents or discussions) are quite similar in the different FOI regimes [3, p. 22].

Katherine Fink made an analysis based on the US FOI Law to answer the question of the accessibility of algorithms. Following her – very logical – structure, first, it is worth to try to answer whether algorithms are a kind of information which is subject to the FOI Law at all and second, to study whether any of the exemptions can be applied to algorithms [9].[8]

Under the Hungarian FOI regime, the subject of the law is "public information", which "shall mean any known fact, data and information, other than personal data, that are processed and/or used by any person or body attending to statutory State or municipal government functions or performing other public duties provided for by the relevant legislation (including those data pertaining to the activities of the given person or body), irrespective of the method or format in which it is recorded, and whether autonomous or part of a compilation."[9] Although the definition is quite wide, it is far not obvious, whether an algorithm used by a public body comes under the scope of this definition.

Even if we accept that the algorithm itself[10] is subject to the FOI Law, still the applicability of some exceptions may arise. First it can be regarded as a "data underlying the decisions" (data for internal use), which may be kept in secret for ten years from the date it was compiled or recorded. It may also be kept in secret after the decision is made, if it supports a future decision as well, or if

[7] Act CXII of 2011 on the Right of Informational Self-Determination and on Freedom of Information (Infotv.)
[8] She also made an empirical research about whether software and algorithms were provided based on a FOI claim, and found also positive results, when e.g. the used software was an open source one, and its operation could be fully analysed.
[9] Infotv., Sec. 3. 5.
[10] At this time I will not deal separately with the legal status of input and output data, which could form the basis for further study later on.

disclosure is likely to jeopardize the legal functioning of the body or the discharging of its duties without any undue influence.[11] It seems that under the Hungarian Law the algorithms may easily be regarded as "data underlying the decision."

Another limitation to access to public information is trade secret. Algorithms of private companies are typically their treasured trade secrets, so their accessibility is based on the company's decision. If the algorithm is developed by the public body itself (or by another public body), the situation is different, since they cannot generally refer to a trade secret as an exception.

Finally, no data is accessible if it is personal data (with some very limited exceptions). The algorithm itself is very unlikely to constitute personal data, but both the input and the output data may be personal data. In this case, data protection regulations shall apply, which exclude general accessibility, but at least grant access rights for the data subject to their own personal data.

4.2. Access to algorithms by the individual concerned

Decisions on individuals may concern natural or legal persons. If the decision of the public body concerns a natural person, than the access rights under the data protection law shall apply. Besides this, both for natural and legal persons access rights to the documents of the proceeding in question (e.g. public administration procedure, civil procedure or criminal procedure) may be applied.

4.2.1. Data Protection Law

The law on data protection is harmonized in Europe, after 25th May 2018, the new European General Data Protection Regulation,[12] the GDPR shall apply.[13] In the Data Protection Law context two definitions are relevant. One is profiling, which "means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements"[14] and the other is "automated decision making", which is not defined generally. Automated decision-making may have two types: (1) the one as defined in Article 22, which refers to "decision based solely on automated processing [without human involvement], including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her", and (2) those automated decisions which do not meet these criteria, whether because they are only partly automated (there are meaningful human interventions), or because they have no legal or similarly significant effect.

Based on the examples summarized above, we can see that in the vast majority of the cases, no fully automated decisions are made, based on profiling, in the public sector, rather human intervention is typical, and the algorithms "only" help the decision maker to decide. Concerning these cases the special rules based on Art. 22 shall not apply. This means that the general data protection rules are applicable supplemented with some special details, based on the interpretation of the general rules.

[11] Infotv., Sec. 27. (5)-(6).
[12] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC
[13] Before that date, the national implementations of the Directive of 95/46/EC (Data Protection Directive) are applicable.
[14] GDPR, Art. 4 (4)

The data subject shall generally have the right to get information before the data processing, and the right of access at any time during the data processing. The obligation covers to provide information about the purposes of the data processing, including the fact that the processing is for the purposes of both profiling and making a decision based on the profile generated; and about the processed data, which means that at least the input data should be made available for the data subject, and some general information about the profile itself. [15] Nevertheless, this does not mean the accessibility of the algorithm itself.

It is not typical (so far) in the public sector that decisions are made without any human intervention, but if the automated decision is solely automated and it has a legal or significant effect on the data subject, more information should be provided. In this case, the data subject has the right to get meaningful information about the logic involved, and about the significance and the envisaged consequences of the decision.[16] According to the Article 29 Working Party, this still does not necessarily mean the disclosure or even the complex explanation of the algorithm, but the data controller "should find simple ways to tell the data subject about the rationale behind, or the criteria relied on in reaching the decision" [2, p. 14].

It is have to be emphasised again that these rules are not applicable to legal persons, even if they are subject to profiling activity and decisions supported by algorithms.

4.2.2. Access to the documents of the proceedings

The second field, applicable both to legal and natural persons is the right to access to the documents of the proceedings in which the person is involved. This field is not harmonized at European level, so the national legislation may vary significantly about the details. Still, it is quite general that the subject of the proceeding has the right to access the relevant documents of the proceeding.[17] This access is far not unlimited, of course. Based on the Hungarian rules, we can see that trade secret for instance may be a reason to limit the access rights of the parties.[18] It is quite common that a software or algorithm, which may help the public body, is a trade secret of a company, or it is under copyright protection. In these cases, the accessibility of the algorithms may be excluded or very limited.

5. Conclusions

The transparency of Big Data and algorithms is and will be a hot topic in legal academic literature, mainly because their relevance is surely going to increase with the evolution of Artificial Intelligence (which is, at this point, mainly based on Big Data and the new analysing methods). In this article, first, the potential usage of these new technologies in the public sector was pointed out, and then we referred to some basic ideas and thoughts about the necessity of making algorithms more transparent.

Then the current legal framework for accessing the algorithms that are used in the course of decision-making by a public body was shown. First, we analysed whether these algorithms are

[15] GDPR, Art. 13, 1., Art. 14. 1., Art. 15. 1. and WP251, p. 23-24.
[16] GDPR, Art. 13. 2. (f), Art. 14. 2. (g), Art. 15. 1. (h)
[17] Cf. in the Hungarian Law, Act CL of 2016 on General Public Administration Procedures (Ákr.) Sec. 33., Act CXXX of 2016 on the Code of Civil Procedure (Pp.) Sec. 162., Act XC of 2017 on the Code of Criminal Procedure (Be.) Sec. 100-102.
[18] Ákr, Sec. 34. Pp. Sec. 163.

accessible under the Freedom of Information regime and second, whether at least those who are affected by the decision have the possibility to have access to the algorithms. The result of this overview has shown that the possibility of access is quite limited in both fields, and it is mainly based upon the decision of the "owner" of the algorithm, or on the decision of the public body. This situation calls for changes: the transparency and accessibility of the algorithms surely has to be improved in the public sector. But the question of "how" will be the subject of another study.

6. References

[1] ALYASS, A., TURCOTTE, M., MEYRE, D., From big data analysis to personalized medicine for all: challenges and opportunities, *BMC Medical Genomics*, 8(33), 2015, https://doi.org/10.1186/s12920-015-0108-y [Accessed 5 December 2017]

[2] ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, Adopted on 3 October 2017 (WP251)

[3] BANISAR, D., *Freedom of Information Around the World 2006*, Privacy International, http://www.freedominfo.org/documents/global_survey2006.pdf [Accessed 25 November 2017]

[4] *Big Data Analysis Proves Effective in Identifying Harmful Drug Interactions*, https://www.dicardiology.com/content/big-data-analysis-proves-effective-identifying-harmful-drug-interactions [Accessed 5 December 2017]

[5] *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, Executive Office of the President, May 2016, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf [Accessed 5 December 2017]

[6] CUKIER, K., MAYER-SCHOENBERGER, V., The Rise of Big Data. How It's Changing the Way We Think About the World, *Foreign Affairs*, 92(3), 2013, pp. 28-40. [Accessed 8 December 2017 via HeinOnline]

[7] DIAKOPOULOS, N., *Algoritmic Acccountability Reporting: On the Investigation of Black Boxes*, Tow Center for Digital Journalism, Columbia University, 2014, http://towcenter.org/wp- content/uploads/2014/02/78524_Tow-Center-Report-WEB-1.pdf [Accessed 5 January 2018]

[8] EPIC, *Algorithms in the Criminal Justice System* https://epic.org/algorithmic-transparency/crim-justice/ [Accessed 4 January 2018]

[9] FINK, K., Opening the government's black boxes: freedom of information and algorithmic accountability, *Information, Communication & Society*, 2017, https://doi.org/10.1080/1369118X.2017.1330418 [Accessed 25 June 2017, via Taylor&Francis Online]

[10] FORGÁCS, I., A Big Data, avagy kezdjünk el gondolkodni a közigazgatás új feladatairól [Big Data, or it is time to start to think about the new tasks of public administration], In: FAZEKAS, M., (ed.): *Gazdaság és közigazgatás: tanulmányok Ficzere Lajos tiszteletére*

[Economy and public administration – essays in honour of Ficzere Lajos], ELTE Eötvös Kiadó, Budapest 2015, pp. 123-133.

[11] GAMAGE, P., New development: Leveraging 'big data' analytics in the public sector, *Public Money & Management*, 36(5), 2016, pp. 385-390. http://dx.doi.org/10.1080/09540962.2016.1194087 [Accessed 25 June 2017, via Taylor&Francis Online]

[12] HAMM, S., *How Big Data Can Boost Weather Forecasting*, Wired, 2013. https://www.wired.com/insights/2013/02/how-big-data-can-boost-weather-forecasting [Accessed 12 December 2017]

[13] HOWARTH, B., *Big data: how predictive analytics is taking over the public sector,* https://www.theguardian.com/technology/2014/jun/13/big-data-how-predictive-analytics-is-taking-over-the-public-sector [Accessed 5 December 2017]

[14] JANSSEN, M., HOVEN, J., Big and Open Data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly*, 32(4). 2015, pp. 363-368. https://doi.org/10.1016/j.giq.2015.11.007 [Accessed 14 September 2017, via ScienceDriect]

[15] LIPTAK, A., Sent to Prison by a Software Program's Secret Algorithms Algorithms in the Criminal Justice System, 1st May, 2017., New York Times, https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html?_r=1 [Accessed 4 January 2018]

[16] MANTELERO, A., Social Control, Transparency, and Participation in a Big Data World, *Journal of Internet Law*, 17(10) 2014 pp. 23-29. http://staff.polito.it/alessandro.mantelero/JIL_0414_Mantelero.pdf [Accessed 25 May 2017]]

[17] MARR, B., *How is Big Data Used in Practice? 10 Use Cases Everyone Must Read* https://www.bernardmarr.com/default.asp?contentID=1076 [Accessed 10 December 2017]

[18] MUNNÉ, R., Big Data in Public Sector. In.: CAVANILLAS, J., M., CURRY, E., WAHLSTER, W. (Eds.), *New Horizons for a Data-Driven Economy. A Roadmap for Usage and Exploitation of Big Data in Europe*, SpringerOpen, 2016. pp. 195-208. https://link.springer.com/content/pdf/10.1007%2F978-3-319-21569-3.pdf [Accessed 5 May 2017]

[19] NAGY, T., Big Data – új (m)érték a büntetőeljárásban [Big Data – new measure in the criminal procedure], In: VÓKÓ, GY., (ed.): *Tanulmányok Polt Péter 60. születésnapja tiszteletére [Essays in honour of the 60th birthday of Polt Péter]*, HVG-ORAC, Budapest 2015, pp. 186-194.

[20] RACSKÓ, P., Big Data a közigazgatásban [Big Data in Public Administration], In: NEMESLAKI, A. (ed.): *E-közszolgálatfejlesztés [Development of E-Public Services]*, Nemzeti Közszolgálati Egyetem, Budapest 2014.

[21] SZŰTS, Z., YOO, J., Big Data, az információs társadalom új paradigmája [Big Data, the new Paradigm of the Information Society], *Információs Társadalom*, 16(1), 2016, pp. 7-28.

[22] *Tackling tax fraud with Big Data approach*, Atos, 2015. https://atos.net/wp-content/uploads/2017/10/atos-risk-and-bi-whitepaper.pdf [Accessed 18 December 2017]

[23] VADÁSZ, P., A Case Study on Finding Fraudulent Practices in the Public Procurement Process Using Text-Mining Methods from Open Internet Sources. In: BALTHASAR, A., et. al. (eds.): *Central and Eastern European e|Dem and e|Gov Days 2016*, Austrian Computer Society, Wien 2016. pp. 471-480.

[24] ZŐDI, ZS., Jog és jogtudomány a Big Data korában [Law and Jurisprudence in the Age of Big Data], *Állam- és Jogtudomány*, 58(1), 2017, pp. 95-114.

[25] ZŐDI, ZS., Privacy és Big Data [Privacy and Big Data], Fundamentum, 2017/1-2. pp. 18-30.