THE PUBLIC DIMENSION OF CYBERSECURITY
M. Karpiuk & J. Kostrubiec

Challenges for State Security in the Context of Big Data Analysis

JUSTYNA KUREK

Abstract The information society is based on constant access to information. The state also performs its tasks with the use of various information databases and information resources of unstructured nature. These resources include information of personal nature, although, notably, personal data is often only a supplementary element, not constituting the main resource being the focus of attention of the state. New tools, such as big data analysis tools, generate additional obligations in the sphere of information security and protection. The author of this paper makes an attempt to identify the potential threats and problems related to the use of big data tools for the processing of information resources of the state, notably, in the context of "incidental" processing of personal data using big data methods. The objective is primarily to draw attention to the specific risks posed by the loss of control over data by the state and the related security implications.

Keywords: • big data • state security • public registers

CORRESPONDENCE ADDRESS: Justyna Kurek, Ph.D., dr. habil., Associate Professor, War Studies University, Faculty of National Security, Department of Political Security, Aleja Generała Antoniego Chruściela „Montera” 103, 00-910 Warsaw, Poland, e-mail: j.kurek@akademia.mil.pl, ORCID: 0000-0002-8754-5243.

https://doi.org/10.4335/2022.1.6 ISBN 978-961-7124-10-1 (PDF)
Available online at http://www.lex-localis.press.

1 Introduction

Today’s information society is built upon having constant, secure access to the information that is necessary both in professional and private life (Yukins, 2004:668). Beyond the search for information, the ability to cope with its overflow and the need to select useful connections becomes a challenge (Vertinsky, Rice, 2002). The state performs its tasks with the use of various information databases and information resources of unstructured nature. These resources contain information of personal nature, although, notably, personal data is often only a supplementary element, not constituting the main resource being the focus of attention of the state. This situation is well illustrated by the examples found within certain national registers run by Polish authorities, e.g. the Register of Entrepreneurs of the National Court Register and the register gathering information about real properties – the centralised Land and Mortgage Register. There is also a noticeable trend in the evolution of the information resources being managed by public institutions towards broadening them by tapping into unstructured resources, which, more and more often, are created through Internet communication. These resources contain personal data processed by the state and its authorities that are often of sensitive nature.

Under these conditions, which are necessitated by *de facto* continuous data analysis, information management mechanisms and technologies, including big data analysis, are of particular importance. These new tools provide effective support in the implementation of tasks in the area of state and national security, generating, however, additional obligations in the sphere of security and protection of information (Kurek, 2021: 122).

The author of this paper makes an attempt to identify the potential threats and problems related to the use of big data tools for the processing of the information resources of the state, in particular, in the context of "incidental" processing of personal data using big data methods. The assumption is primarily to draw attention to the specific risks posed by the loss of control over data by the state and the related security implications.

2 Big data phenomenon and big data analysis methods

The big data phenomenon is often described as the adoption of new technologies or the application of a set of new technical tools that facilitate data collection and mathematical analysis using traditional statistical methods, as well as more innovative analytical approaches. However, the source literature notes that this view may not fully capture the nature of the phenomenon, especially its power and uniqueness (Mayer-Schönberger, Padovao, 2016: 318). Big data opens up a new perspective on reality. V. Mayer-Schönberger and K. Cukier figuratively define big data processes as "enabling data to speak" (Mayer-Schönberger, Cukier, 2014: 9), while I.S. Rubinstein perceives the big data phenomenon in terms of "steroidally" stimulated data mining processes (Rubinstein, 2013: 76). There is no consensus among legal commentators on the definition and key characteristics of this phenomenon. Certainly, it should be evaluated dynamically because technological development and new applications significantly affect its understanding and distinguish it from other forms of data analysis (Broeders, Schrijvers, Sloot, Brakel, Hoog, Hirsch, 2017: 310).

It is common to draw attention to three defining qualities figuratively referred to by English language legal commentators as the 3Vs – Volume, Variety and Velocity (Klous, Sustainable, 2016: 27-47). Big data volumes are thus characterised by three basic qualities. The first one is mass availability of data – collected not only from online sources, but also through mobile devices equipped with localisation services and numerous data distribution applications, as well as information from objects equipped with artificial intelligence (Internet of Things) (Hildebrandt 2012: 45-46). The second quality is the use of high speed processing devices and data transfer to achieve cheap and efficient data processing. This analysis more and more often additionally uses the cloud computing model. The third quality is the use of new computing frameworks to collect and analyse massive volumes of data (Rubinstein: 2013: 74). This model can be further complemented by a fourth V (Value) referring to data value (Szafranski, 2015: 11).

Big data processes have undoubtedly changed the face of data analysis, certainly representing a new model of information management in both business and organisational aspects. Indeed, data can be reused for purposes other than the purpose of its original collection. Moreover, data value can be increased not only through new collection and analysis processes, but also by linking data with data from other sources (Kurek, 2021: 126). Data mining also facilitates discovery or inference of previously unknown facts and patterns from the database.

While in the traditional view, data value was manifested in its collection and single use for a specific purpose, big data processes have introduced a revolution, according to which the informational value of data is unclear at the time of their collection (Mayer-Schönberger, Padovao, 2016: 319).

3 Databases in the service of state security

The state and its authorities are the keepers of numerous databases, and most public registers are kept in this form. As M. Kiedrowicz noted in his research, in 2015, according to various sources, the number of registers and records kept in Poland, ranged from 600 to 3000. The scope of information that is collected, stored, processed and further made available by them is vast (Kiedrowicz, 2015: 30). However, this is mostly structured data. It is noted, however, that only 15% of all information produced by humanity is structured and suitable for processing using relational database methods and tools. The remaining 85% constitutes a large 'reservoir of data', whose informational content is undoubtedly invaluable, but due to its unstructured nature, is unsuitable for processing in an organised manner (Dygaszewicz, 2015: 49). Its re-analysis and use by public authorities is only possible due to the potential of big data technology, which facilitates re-organisation and re-analysis of resources for the purpose of obtaining information, the potential of which was not originally envisaged.

The use of big data analysis for the performance of the tasks in the area of state and national security is a major challenge currently faced by state authorities. These challenges are of both organisational and legal nature. Having regard to the principle of legalism, the state authorities may act only within and under the law, therefore, without an appropriate legal basis; they may not process data for purposes other than those for which they were collected. Moreover, the very process of data collection requires an adequate legal basis. Taking into account the fact that in the case of big data processes, the purpose of data use is *de facto* not known at the moment the data comes into possession, the processing of big data may pose particular challenges for public authorities (Kurek, 2021: 138).

It is, therefore, difficult to organise the protection of information in a preventive manner if the way it will be used and linked to other data is not fully known. A key element of data and information management policy at the initial stage of the legislator's decision to create a relevant resource and database (in particular, one that contains personal information) should be proper risk analysis. Such analysis should include both an in-depth reflection on the processes connected with processing, safety of collection and sharing of data, but also on the security of data sets, so that in case of losing control over given data, it cannot be easily used or manipulated.

Structured data aggregated into a relational database pose a huge challenge in the sphere of security. It is insufficient from the security point of view to concentrate only on the external layer and on securing only entry to the system. Breaching the external security protecting against all forms of unauthorised access may be just a matter of proper combination of queries to the database and de facto be a security bypass, not a security breach. A perfect example is the bypass of security protecting the land and mortgage register resource, which took place several years ago. The only real security of this system protecting against an automatic takeover of the resource by means of automatic queries is the CAPTCHA mechanism, which *de facto* does not generate protection against automated access (Ahn, Blum, Langfords, 2004: 57-60).

4 The *casus* of re-use of land and mortgage register data

The risks for state security and privacy are perfectly illustrated by the example of re-use of the data in the Land and Mortgage Register. An entity having its registered office in the Seychelles collected and indexed information from over twenty million land and mortgage registers (Gryszczyńska, 2017: 298). This procedure was possible even though, theoretically, public access to land and mortgage registers is possible only through one search criterion – the land and mortgage register number. The collection of the specified resource was not the outcome of obtaining the unique numbers of over twenty million land and mortgage registers, nor of breaking the security measures and obtaining data in an illegal manner. The entity that collected the specified information did so by working out how the land and mortgage register numbers were constructed.

It should be noted that the register number consists of three predefined elements: a court district code (to be selected from a list), a specific number consisting solely of digits and a checksum between 0 and 9. Taking into account the limited number of specific numbers and knowledge of the two additional elements of the register number, created for the purpose of database queries and obtaining information, a list of potential numbers of land and mortgage registers was easy to generate. It was therefore relatively easy to extract the structured data and re-enter it into databases managed by another entity and to apply additional search criteria (e.g. real property address, plot registration number, owner's name, existence or not of mortgage encumbrances). This way, through the re-use of public information, it was possible to build a system facilitating extraction of information about owners of specific real properties or to simply obtain information about mortgage encumbrances and the amount of loans with which a given real property has been financed. This information could be easily used for criminal purposes and might be an excellent source of information for criminals, as noted in available studies on this subject (https://www.rp.pl/artykul/988227-Ksiegi-wieczyste--wyciekly-dane-o--16-milionach-hipotek-w-Polsce.html).

This generates not only the risk for the privacy and security of specific individuals, but also for the state and its authorities, which, as one of their key objectives, ensure security to its citizens, as well as all persons and property on their territory. Land and mortgage registers contain information on the property and possessions of key people in the state. The address data provided in the system also facilitate a potential identification of the place of residence of the key persons in the state.

Extracting the information in question was not the outcome of a criminal offence or a breach of security, nor was it the outcome of unlawful entry into possession of state-managed information. It was the outcome of security bypass and re-use of public information in accordance with the law in force. The structuring of the data only facilitated the reprocessing. One could wonder whether in this situation it is possible to speak of an abuse of right in the meaning of Article 5 of the Civil Code, i.e. the use of a subjective right (the right to re-use public information) contrary to its socio-economic purpose or principles of community life. In my opinion, such interpretation is too far-reaching and *de facto* annihilates the political objective of the institution of re-use of public information.

Of course, the question should be asked whether meeting the objective of openness of land and mortgage registers required such a form of access to data and their full centralisation, as it was done by the Polish legislator, who decided to fully digitalise and centralise public registers. It is worth mentioning the examples from, for instance, Germany, where obtaining an extract from the land and mortgage register is done through the portal of justice (www.justiz-portal.de). In Germany, data sets were not centralised and there is not just one database. The website, referred to above, only contains links to the portals maintained in individual states (Länder). Access groups and gradation of access rights have also been introduced. For example, unconditional and full access is granted to notaries and real estate institutions (insurers, banking institutions, administrative offices, courts). For others, access is possible but only upon fulfilment of access conditions and, in some cases, upon payment of a symbolic fee.

Perhaps in Poland we should also think of decentralisation of registered data sets, by way of consolidation of the same through common links. It is also worth considering whether the information should not be managed in the form of a database system or if it would be sufficient to make it available in a form aggregated to a closed pdf format with protection against copying. One could also set a question if, from a security point of view, the procedure for numbering of the register should not be re-established, so that they are numbered at random rather than according to a template. Perhaps the difficulty of working with such a system and managing such data would not outweigh the gain in information security.

It is also worth asking if at least some of the personal data included in the public resource should not be anonymised or hidden. From the point of view of security of the conduct of legal transactions, information that is truly important is the mortgage collateral, but the information about the value of the collateral could be available only to entities having legal interest in obtaining such information. Indeed, from the point of view of state security, security measures and access levels may play two functions: on the one hand, they facilitate the control of information managed by the state, on the other hand, they introduce the control of access and make it possible to record the recipients of information.

Public access to data in the land and mortgage register also implies the use, in the conduct of legal transactions (with the legislator’s consent), of extracts from the register made individually in an unauthorised manner. The practice is that extracts from the register are made personally by the parties to a legal transaction and attached to the documentation. When unauthorised sets of information are created using reprocessed public information, there is also a risk that extracts from such private databases will be made and submitted instead of extracts from public registers. In the case of such private database systems, the consolidation of information and its accuracy is not covered by the public quality guarantee in the form of, in the case at hand, the warranty of public credibility of land registers.

5 Conclusion

As perfectly illustrated by the example of the processing of land and mortgage register data in the Seychelles, unauthorised re-use of personal data by a data controller reveals a completely new potential of data abuse. This often implies a serious security risk for persons whose data is – even incidentally – processed. Therefore, one should ask if the potential of using big data tools outweighs the threats and challenges for state security posed by consolidation and integration of data of various provenance in terms of big data. In practice, loss of control over data and security threats do not necessarily result from illegal access to information. They are often the result of lawful use of public information in the mode of re-use of public information for a purpose other than the purpose of its extraction. Hence, when it comes to information management, particularly in the era of big data analysis, risk analysis is crucial. This applies across the board to those processes where the data potential is unknown to the data controller at the outset. Effective information security management requires preventive measures, and, just as in war, the greatest success is to defeat the enemy without a fight, so in the case of information security the most important issue is to effectively predict and counteract the risks. The revealed loopholes in the system, which result from legal regulations, on the one hand, and from the possibility to implement them, on the other hand, indicate that the lack of risk analysis and adequate data and information processing security may result in the risk of losing control over data exceeding the advantages related to the potential of big data and data consolidation from various resources.

References:

Ahn, L., Blum, M. & Langfords, J. (2004) Telling Humans and Computers Apart Automatically, *Communications of the ACM*, 47(2), pp. 56-60.

Broeders, D., Schrijvers, E., Sloot, B., Brakel, R., Hoog, J. & Hirsch, E. (2017) Big Data and Security Policies: Towards a Framework for Regulating the Phases of Analytics and Use of Big Data, *Computer Law & Security Review*, 33(3), pp. 308-323.

Dygaszewicz, J. (2015) Big data w statystyce publicznej, In: Szpor, G. (ed.) *Internet. Publiczne bazy danych i Big Data* (Warszawa: C.H.Beck), pp. 49-63.

Gryszczyńska, A. (2017) Nowe zagrożenia rejestru ksiąg wieczystych, In: Szpor, G. & Gryszczyńska, A. (eds.) *Internet. Strategie bezpieczeństwa* (Warsaw: C.H.Beck), pp. 293-310.

Hildebrandt, M. (2012) The Dawn of a Critical Transparency Right for the Profiling Era, In: Bus, J., Crompton, M., Hildebrandt, M. & Metakides, G. (eds.) *Digital Enlightenment Yearbook* (Amsterdam: IOS Press), pp. 41-56.

Kiedrowicz, M. (2015) Dostęp do publicznych zasobów danych. Big data czy big brother, In: Szpor, G. (ed.) *Internet. Publiczne bazy danych i Big Data* (Warszawa: C.H.Beck), pp. 15-41.

Kurek, J. (2021) *Bezpieczeństwo państwa w warunkach hybrydowej regulacji danych osobowych w dobie analizy Big data. Aspekty prawne, organizacyjne i systemowe* (Warszawa: ASzWoj).

Mayer-Schönberger, V. & Cukier, K. (2014) *Learning with big data* (Boston-NewYork: Houghton Mifflin Harcourt Publishing Company).

Mayer-Schönberger, V. & Padovao, Y. (2016) Regime Change? Enabling Big Data through Europe's New Data Protection Regulation, *The Columbia Science and Technology Law Review*, 17(2), pp. 317-334.

Rubinstein, I.S. (2013) Big Data: The End of Privacy or a New Beginning?, *International Data Privacy Law*, 3(2), pp. 74-87.

Szafrański, B. (2015) Realizacja zadań publicznych a Big data, In: Szpor, G. (ed.) *Internet. Public databases and Big Data* (Warszawa: C.H.Beck), pp. 3-15.

Vertinsky, L. & Rice, T.M.. (2002) Thinking about Thinking Machines: Implications of Machine Inventors for Patent Law, *Boston University Journal of Science and Technology Law*, 2, pp. 574-613.

Yukins, C.R. (2004) Making Federal Information Technology Accessible: A Case Study in Social Policy and Procurement, *Public Contract Law Journal*, 33(4), pp. 667-725.