USE OF OFFICIAL DATA OF STATE INSTITUTIONS IN THE SCIENTIFIC RESEARCH OF THE PAPULATION SECURITY[1]

Vidmantas Egidijus Kurapka

Mykolas Romeris University, Faculty of Law,
Department of Criminal Procedure
Ateities 20, LT-08303 Vilnius, Lithuania
Phone (+370 5) 2714 533
E-mail egidijus@mruni.eu

Viktoras Justickis

Mykolas Romeris University, Faculty of Social Police,
Department of Psychology
Ateities 20, LT-08303 Vilnius, Lithuania
Phone (+370 6) 1533 289
E-mail justickv@takas.lt

Abstract. The paper discusses the problems in the detection of security information in legal and other administrative data. The authors analyse the prospects of the use of data-mining in the solution of two key problems: abundance and indirectness of these data. Security research uses two kinds of data. The first one is scientific data, designed and gathered specially for the verification of certain security theories. They are the data of criminological,

---

---

sociological, psychological surveys, experimental data, etc. The second kind is data that are not designed for security research. Most of the data are the by-products of the national legal system, especially its criminal justice, of national and local institutions responsible for the maintenance of public order, and of other public agencies. All the data provide huge amounts of information destined to control and direct the activities of these institutions. There are two kinds of problems in the use of this information in security study. First, this information has to be 'decoded' from data describing the activities of related institutions. Second, data on security are lost in huge amounts of other public data. Thus, they have to be 'mined'. The paper discusses the prospects of the modern information proceeding method—data-mining in the solution of both problems. A new concept of data mining as a meta-procedure is proposed. The object of this meta-procedure is supposed to be the integration of multiple current theories (for example, criminal security) and related statistical procedures. A general algorithm of such data-mining is proposed.

Keywords: *security, criminality, criminal security, data-mining.*

Introduction

'Security' has become the core concept integrating criminological, sociological, economical, medical, ecological, etc. studies of society.[2] Over the last several decades, the term 'security' has become highly popular, especially in social sciences. Law, sociology, ecology, political science, medicine increasingly discuss legal, social, ecological, political, medical securities of the world, its parts, and single countries. Every branch of science is concerned with its own security problems, powers, trends, accidents, catastrophes. Everyone shows ways in which a danger for the society, its groups, and its single citizens can arise.

Criminal security (security from becoming a victim of a crime and experiencing any harm due to criminal activities) is one of the most prospective directions of security studies.[3] Also, as other directions, it provided a mighty impetus for a large scale integration of quite different disciplines for criminal security research.[4]

Legal sciences, especially criminal law, criminology and a new quickly developing branch of modern criminalistics, i.e. 'preventive criminalistics', discuss risks coming

2   Kurapka, V. E., *et al*. Security and Economic Development. Interaction within the European Union: An Integration Model. In *Third Year Within the European Union: Topical Problems in Management of Economics and Law*. Judzins, T.; Justickis, V. (eds.). Riga: Latvian Academy of Science, 2007, p. 163−179.

3   Kurapka, E.; Malevski, H.; Justickis, V. Das staatliche Wissenschaftsprogramm "Dynamik, Kontrolle und Prognosen der Kriminalität in Litauen. Die gegenwärtige Konzeption der Kriminalistik." Anpassung des litauischen Rechtssystems an die EU-Standards. In *Die Umsetzung des Gemeinschaftsrechts im innerstaatlichen Rechts*. Koeck, H.; Banaszak, B.; Kurapka, E. (eds.). Lindz: Lindz University, 2006, p. 259−270.

4   Kurapka, E.; Malevski, H.; Justickis, V. Strategische Richtungen der Kriminalitätsprävention und deren Harmonisierung im sich vereinenden Europa. In *Die Umsetzung des Gemeinschaftsrechts im innerstaatlichen Rechts*. Koeck, H.; Banaszak, B.; Kurapka, E. (eds.). Lindz: Lindz University, 2006, p. 251−258.

from dangerous developments in criminality.[5] Sociology discusses criminal dangers coming from social conflicts, dysfunctions of the social structure, deviant and destructive behaviour of individual persons or social groups.[6] Political sciences analyse political risks stemming from political activities, especially illegal ones.[7] Ecological and medical sciences review criminal threats coming from natural processes and catastrophes: natural disasters, epidemics, etc.[8] Economics provides its own insights into criminal threats arising from the current economic trends.[9]

All this attracts and uses huge amounts of data, different in their source, nature, design. This brings new challenges for data processing. The ability to go ahead in security research is now highly dependent on the ability to collect and process related data.

Security research uses two kinds of data. The first one is scientific data, designed and gathered specially for the verification of certain security theories. They are the data of criminological, sociological, psychological surveys, experimental data, etc. The second kind is data that are not intended for security or any other research. Most of the data are the by-products of the national legal system, especially its criminal justice, of national and local institutions responsible for the maintenance of public order, and of other public agencies. All the data provide huge amounts of information destined to control and direct the activities of these institutions.

The main problem in the use of these data for security studies is that they have not been designed to analyse criminal security and factors influencing it. The aim of all these data is quite different—to ensure smooth functioning of the abovementioned institutions and the whole criminal justice. Therefore, when a security researcher intends to use these data, he meets several problems. The most important of them are two.

The first one is the abundance of information. Probable data on security are lost in huge amounts of other public data. This creates a paradoxical situation: the profusion of information on any research subject is usually seen as a highly favourable situation, but, contrary to that, in criminal security studies, this abundance of information becomes an obstacle. Criminal security information has to be 'mined' in lots of other data.[10]

The second problem is that most of information found in these data does not describe security directly. They are intended to represent the activities of related institutions. Most often these data reveal the security situation only indirectly, in the light of the

5   Kurapka, V. E., *et al.* Dynamics and Prognosis of the Criminality in Lithuania and Modern Conception of Criminalistics. In *Politics and Law in the Context of European Integration*. Judzins, T.; Justickis, V. (eds.). Riga: LZA BSPC, 2003, p. 106−110.

6   Pečkaitis, J.; Justickis, V. Development of the Policing Concept in the Transitional Lithuania. In *Menschenrechte und öffentliche Sicherheit*. Meikališa, A.; Justickis, V.; Kavalieris, A. (eds.). Riga: Polizeiakademie Lettlands*,* 2004, p. 173−193.

7   Sears, D. O.; Huddy, L.; Jervis, R. (eds.). *Oxford Handbook of Political Psychology.* New York: Oxford University Press, 2003.

8   Pirages, D.; Ken Cousins, K. (eds.). *From Resource Scarcity to Ecological Security. Exploring New Limits to Growth*. Cambridge: MIT Press, 2005.

9   Snowdon, B.; Howard, R. *Modern Macroeconomics: Its Origins, Development and Current State*. Cheltenham: Edward Elgar Publishing, 2005.

10   Xu, J.; Chen, H. Criminal network analysis and visualization. *Communications of the ACM*. 2005, 48(6): 100−107.

objects and activities of these institutions. Therefore, any criminal security study faces the challenge to discern the criminal security information in large and super-large collections of data, to 'decode' it.

The modern data-mining (DM) is an information processing method designed to 'mine' such 'masked', latent information hidden in great amounts of other data. Our paper discusses the prospects of the use of DM in such criminal security studies that use institutional (non-scientific) data.[11]

The paper focuses on the integration of the modern criminological theory and statistical data mining methods for the detection of criminal security information in large and super-large collections of data. A new concept of DM as a meta-procedure for the integration of criminological theory and statistical methods in security research is discussed.

The paper consists of five parts. In the following part, the authors discuss the modern concepts of data-mining and show their inconsistency. The third part delineates the set of concepts related to criminal data and criminal theory. The fourth part deals with special demands to DM stemming from the specific nature of crime data. This provides the basis for the development of an algorithm supposed to combine criminal theories and DM for the discovery and analysis of crime and crime security data. The fifth part includes considerations regarding the future criminal security studies.

1. The Concept of Data-Mining. Data-Mining as a Meta-Procedure

The modern use of the term 'data-analysis' is rather amorphous.[12] Most often it is associated with the processing of large amounts of information, discovering 'hidden information' in these data, use of rather sophisticated, 'non-trivial' statistical procedures, interactive data processing.[13]

All these definitions are based on the idea that DM represents a special group of statistical procedures. This approach is not correct. In fact, all known statistical procedures are used in DM. Also, there exist no exclusively DM procedures.

We propose a different concept of DM. This concept is supposed to delineate the specifics of DM and its place in the processing of data.

We suppose that the DM *is not* the name of any special group of statistical procedures (for example, of sophisticated, 'non-trivial', etc). Instead, DM is *a meta-procedure* of data processing, which means a 'procedure for the application of other procedures'. DM is a *procedure showing how different statistical procedures should be used when*

---

11   McCue, C. *Using Data Mining to Predict and Prevent Violent Crimes* [interactive]. [accessed 22-08-09]. <http://www.spss.com/dirvideo/richmond.htm?source=dmp age&zone=rtsidebar>.

12   Giudici, P. *Applied Data Mining: Statistical Methods for Business and Industry*. New York: John Wiley & Sons, 2003.

13   Raudys, S. *Statistical and Neural Classifiers: an Integrated Approach to Design*. London: Springer, 2001; Giudici, P., *op. cit*.

exploring a certain kind of data. Thus, DM is *not* the use of trivial or sophisticated procedures, *not* the collection of intuitive or rigorous ones, *not* any group of procedures used to process multiple or abundant data. Instead, DM is a meta-procedure showing how these quite different statistical procedures have to be applied when analysing a certain sort of data.

In the case of crime data, DM procedure indicates when and which data processing procedure should be applied in the analysis. DM procedure is supposed to show, for example, on which stage, in what succession, in which situation, etc. of criminal data processing single trivial or sophisticated, intuitive or rigorous, etc. statistical procedures should be applied.[14]

DM is a new concept which appeared as an answer to new challenges. DM was not so necessary 30-40 years ago. At that time, statistical procedures were not very numerous and were applied to more homogenous and specified data. When analysing separate kinds of data, a common experience of the analyst was quite sufficient to select and adapt statistical procedures to the nature of data. The selection of a proper set of tools and the way of their usage was a matter of the analyst's individual art and skill.[15]

Contrary to this, today a boundless variety of different statistical tools and methods is available. Each of them has its own prevalence and restrictions, its own destination and particular features. Furthermore, data became far more varying. Of utmost importance is the great increase in a very special kind of data—social data—characterized by a great degree of vagueness and uncertainty, highly intuitive nature, ambiguity about their interpretations, etc.

Under such new circumstances, the selection of suitable analytical tools became really complicated. Therefore, general meta-procedures facilitating such a selection are highly necessary. All this implies four fundamental requirements for the procedure of DM:[16]

1. Both the selection of statistical tools and the consistency of their usage have to be dependant on the specific nature of data.

2. This specific nature is described by related science and its theory. The latter shows both the peculiarity of its data and their interconnection—the 'structure of data'.

3. It means that distinctive requirements for DM of certain data have to be inferred from the theory describing these data.

4. This, in turn, means that these distinctive requirements only must determine the selection of statistical tools, the order and way in which they are used while processing the related data.

---

14   Raudys, S., *et al*. Functional model of criminality: simulation study. *Lecture Notes in Artificial Intelligence*. 2005, 3554: 410−423.

15   Kloesgen, W.; Zytkow, J. (eds). *Handbook of Data Mining and Knowledge Discovery.* New York: Oxford University Press, 2002.

16   Justickis, V. *Criminology*. I-II. Vilnius: Mykolo Romerio universitetas, 2001, 2004, p. 956.

2. Crime Theories and Their Usage in the Processing of Crime Security Data

Criminology is a discipline responsible for the description of criminal behaviour, discovering its reasons and in this way explaining it. The concept 'crime data' embraces, first of all, the records of the police and courts, correction and other institutions liable for crime control. Criminology describes and explains criminal behaviour and its reflection—criminal data—by using *criminological theories*. As shown, requirements for the DM of specific data should be inferred from the theory describing these data. Let us review the distinctive features of criminological theory. It should be stated right away that criminological theory is very different from the theories of many other sciences, especially from the theories of the so-called exact sciences.

1. *Plurality of crime theories.* The most usual situation in the great majority of disciplines is the recognition of only one theory as correct. A new theory replaces the old one. As opposite, a normal situation in criminology is the plurality of crime theories, all of them are considered to be correct. Fundamental handbooks of criminology describe more than a hundred of modern criminological theories. None of them has any recognized prevalence, none is considered to be more correct than the others.

2. *Hypothetical status of crime theories.* In the great majority of disciplines, theories contain statements that are used without any new verification. For example, after the gravitation theory was substantiated, there is no need to re-check it every time when dealing with the manifestations of gravitation.

The situation in criminology is quite different. When dealing with criminal behaviour (for example, when searching for the reasons of street criminality), *one cannot say in advance which of the many modern criminological theories will be applicable to explain it.* Instead, different theories propose different (hypothetical) explanations; each of them indicates a *different set of factors* supposed to cause the criminal behaviour in question. It means that it is up to the researcher to review all possible explanations and to find the proper one.

3. *Noisy context of the operation of criminological theories.* Criminal behaviour operates in a context permeated by values, emotions, interests. Both criminal acts and social reaction to them are highly changeable depending on myriads of influences, events, effects, often of incidental nature. Therefore, the manifestations of factors described by every theory can be extremely different.

All this is the reason why the interconnection between these factors and crime data representing them is highly variable.

4. *Problematical nature of causal statements in criminology.* The basis of every scientific theory is statements about the causal interaction between certain events. However, the verification of such statements is the most problematic in criminology. The point is that the criminological theory describes the interaction of objects which usually cannot be experimentally manipulated: society, social groups. Therefore, causal statements are verified indirectly, using time observations and gathering additional data.

3. Nature of Crime Data and Data-Mining

There are specific demands to criminal DM stemming from the specific nature of crime data:

1. DM should be based on the bank of criminological theories that can be used for the study of given data. Ideally, this bank should contain all modern criminological theories, because each of them can provide its own assumption on the probable structure of data.

2. DM should contain series of procedures for the modification of data. They are series of general modifications applied to all data (calculation of averages, correlations, etc.) and series of the specific ones applied to a single theory.

3. This bank has to include a serial description of the results of these modifications—probable manifestations of every theory in the data. The serial description of a single theory indicates its probable manifestations based on every statistical modification of data.

4. The algorithm should be built on the interaction of two general super-blocks. The first one is the general overview of all data. Its aim is to discover the manifestation of some theories included into the theory bank and to single out the most promising of them. It is based on general modifications. The second is the special one. It is DM for the most promising theory. On this stage all data are studied as well. However, it is done for a different aim—to get additional information to support (or deny) the most promising theory. If the probability that a certain theory 'works' becomes high enough, this theory is included into the bank of final explanations.

5. Theories found to be not promising or theories the manifestations of which were not observed or were too weak are not excluded from the theory bank. This means that their manifestations will be searched for in every new general modification of all data.

6. The interaction between general and specific super-blocks continues until the series of general modifications is ended. The outcome of this interaction can be a small group of theories that will proceed to the final block-verification of their causal statements. The flow block of criminal DM algorithm that realize these general demands are presented in Figure.

4. Considerations for Future Criminal Security Studies

The outlined crime DM algorithm aims to transform crime data analysis from the present magic and black art to a well-ordered and transparent process. Further efforts of criminologists and DM experts designing this algorithm should proceed in the main directions outlined in this report: specification of the list of criminological theories to be included into the theory bank, development of the series of data modifications and the description series for every criminological theory included into theory bank. This, in turn, will provide the basis for the proposed algorithm of crime DM and the prospect for its use and investigation.
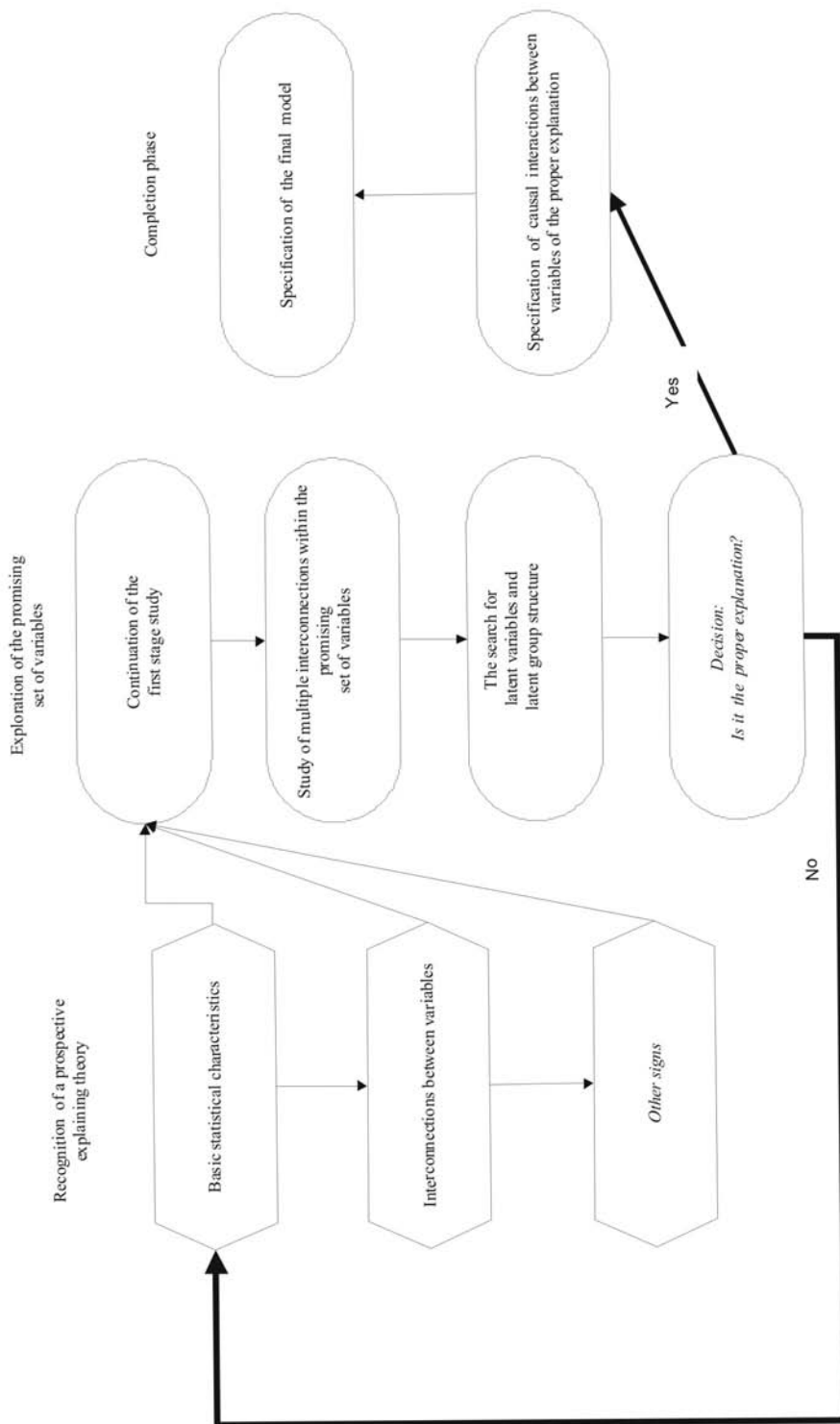
The data-mining algorithm for crime data

Conclusions

1. The greatest part of the modern criminal security research is based on non-scientific data, which means that they are not designed for the verification of security statements. Institutional data (especially produced by criminal justice institutions) are used most intensively in the studies of criminality). The use of these data provides important additional problems caused by the abundance and indirectness of these data.

2. The modern DM is a promising solution of both problems. In the study of a certain object, it integrates scientific theories dealing with this object with the related statistical procedures.

3. The concept of DM used in the current literature is rather amorphous and should be specified. The most promising way for such a specification is the view of DM as a meta-procedure. According to this view, DM is considered to be a 'procedure indicating how to use statistical and other information processing procedures'. This means that the basis of DM is a broad algorithm-indicated way in which consequent theories are tried to be adapted to specific data, related statistical procedures are selected and used, the outcomes of this usage are interpreted and the direction for further research is chosen.

4. The design of crime security DM algorithm has to be based on a bank of criminological theories that can be used for the study of given data.

5. Criminal security DM algorithm should contain a series of procedures for the modification of data. They are the series of general modifications applied to all data (calculation of averages, correlations, etc.) and the series of the specific ones applied to separate theories. This bank should include a serial description of the results of these modifications—probable manifestations of every theory in the data. The serial description of a single theory indicates its probable manifestations depending on every statistical modification of data.

6. The algorithm should be built on the interaction of two general super-blocks. The first one is the general overview of all data. Its aim is to discover the manifestation of some theories included into the theory bank and to single out the most promising of them. It is based on general modifications. The second is the special one. It is DM for the most promising theory. On this stage all data are studied as well. However, it is done for a different aim—to get additional information to support (or deny) the most promising theory. If the probability that a certain theory 'works' becomes high enough, this theory is included into the bank of final explanations.

7. Theories found to be not promising or theories the manifestations of which were not observed or were too weak are not excluded from the theory bank. This means that their manifestations will be searched for in every new general modification of all data.

8. The interaction between general and specific super-blocks continues until the series of general modifications is ended. The outcome of this interaction can be a small group of theories that will proceed to the final block-verification of their causal statements.

References

Giudici, P. *Applied Data Mining: Statistical Methods for Business and Industry*. New York: John Wiley & Sons, 2003.

Justickis, V. *Criminology*. I-II. Vilnius: Mykolo Romerio universitetas, 2001, 2004.

Kloesgen, W.; Zytkow, J. (eds). *Handbook of Data Mining and Knowledge Discovery*. New York: Oxford University Press, 2002.

Kurapka, E.; Malevski, H.; Justickis, V. Das staatliche Wissenschaftsprogramm "Dynamik, Kontrolle und Prognosen der Kriminalität in Litauen. Die gegenwärtige Konzeption der Kriminalistik." Anpassung des litauischen Rechtssystems an die EU-Standards. In *Die Umsetzung des Gemeinschaftsrechts im innerstaatlichen Rechts*. Koeck, H.; Banaszak, B.; Kurapka, E. (eds.). Lindz: Lindz University, 2006.

Kurapka, E.; Malevski, H.; Justickis, V. Strategische Richtungen der Kriminalitätsprävention und deren Harmonisierung im sich vereinenden Europa. In *Die Umsetzung des Gemeinschaftsrechts im innerstaatlichen Rechts*. Koeck, H.; Banaszak, B.; Kurapka, E. (eds.). Lindz: Lindz University, 2006.

Kurapka, V. E., *et al*. Dynamics and Prognosis of the Criminality in Lithuania and Modern Conception of Criminalistics. In *Politics and Law in the Context of European Integration*. Judzins, T.; Justickis, V. (eds.). Riga: LZA BSPC, 2003.

Kurapka, V. E., *et al*. Security and Economic Development. Interaction within the European Union: An Integration Model. In *Third Year Within the European Union: Topical Problems in Management of Economics and Law*. Judzins, T.; Justickis, V. (eds.). Riga: Latvian Academy of Science, 2007.

McCue, C. *Using Data Mining to Predict and Prevent Violent Crimes* [interactive]. [accessed 22-08-2009]. <http://www.spss.com/dirvideo/richmond.htm?source=dmpage&zone=rtsidebar>.

Pečkaitis, J.; Justickis, V. Development of the Policing Concept in the Transitional Lithuania. In *Menschenrechte und öffentliche Sicherheit*. Meikališa, A.; Justickis, V.; Kavalieris, A. (eds.). Riga: Polizeiakademie Lettlands*,* 2004.

Pirages, D.; Ken Cousins, K. (eds.). *From Resource Scarcity to Ecological Security. Exploring New Limits to Growth*. Cambridge: MIT Press, 2005.

Raudys, S. *Statistical and Neural Classifiers: an Integrated Approach to Design*. London: Springer, 2001.

Raudys, S., *et al*. Functional model of criminality: simulation study. *Lecture Notes in Artificial Intelligence*. 2005, 3554: 410−423.

Sears, D. O.; Huddy, L.; Jervis, R. (eds.). *Oxford Handbook of Political Psychology*. New York: Oxford University Press, 2003.

Snowdon, B.; Howard, R. *Modern Macroeconomics: Its Origins, Development and Current State*. Cheltenham: Edward Elgar Publishing, 2005.

Xu, J.; Chen, H. Criminal network analysis and visualization. *Communications of the ACM*. 2005, 48(6): 100−107.

VALSTYBĖS INSTITUCIJŲ OFICIALIOSIOS INFORMACIJOS PANAUDOJIMAS GYVENTOJŲ SAUGUMO MOKSLINIUOSE TYRIMUOSE

Vidmantas Egidijus Kurapka, Viktoras Justickis

Mykolo Romerio universitetas, Lietuva

Santrauka. *Straipsnyje nagrinėjama, kaip valstybės institucijų (ypač baudžiamųjų justicijų) duomenys panaudojami saugumo tyrimams. Skirtingai negu mokslinė informacija, kuri specialiai gaunama tam tikriems moksliniams teiginiams patikrinti, institucinė informacija yra šalutinis šių institucijų veiklos produktas. Ji nėra tiesiogiai skirta saugumo problemoms tirti. Jos paskirtis yra užtikrinti atitinkamos institucijos veiklos kontrolę ir valdymą. Dėl to atliekant saugumo tyrimą panaudojant šiuos duomenis susiduriama su nelengvomis papildomomis problemomis. Dvi iš šių problemų yra itin svarbios.*

Pirma, duomenys, kuriuos teikia valstybės institucijos, dažniausiai tik netiesiogiai atspindi saugumo padėtį. Tiesiogiai šie duomenys parodo tik tai, kas domina šias institucijas ir taip, kaip to reikia jų veikloje. Saugumo tyrėjas turi rasti būdą „atkoduoti" esančią juose informaciją apie saugumą bei jį lemiančius veiksnius.

Antra vertus, duomenys, kurie gali teikti tam tikrą informaciją apie saugumą, yra paslėpti didžiuliuose įvairiausios institucinės informacijos kloduose. Tyrėjas dar turi surasti tuos duomenis, nustatyti, kokie įvairiausi instituciniai duomenys gali būti naudingi apibūdinant saugumo padėtį ir jo veiksnius. Straipsnyje nagrinėjamos dabartinio duomenų paieškos (data-mining) metodo galimybės, sprendžiant minėtas problemas.

Siūloma nauja duomenų paieškos koncepcija. Pagal ją duomenų paieškos metodas yra metaprocedūra, skirta suderinti statistinių metodų panaudojimą tiriant tam tikrą objektą su tą objektą apibūdinančiomis teorijomis. Aptariama, kaip taip suprantama duomenų paieškos metaprocedūra gali būti taikoma vykdant kriminalinio saugumo tyrimus. Siūlomas atitinkamas duomenų paieškos algoritmas, skirtas naudoti kriminalinio saugumo tyrime.

Reikšminiai žodžiai: *saugumas, saugumo tyrimas, instituciniai duomenys, duomenų paieška.*

Vidmantas Egidijus Kurapka, Mykolo Romerio universiteto Teisės fakulteto Baudžiamojo proceso katedros profesorius. Mokslinių tyrimų kryptys: baudžiamoji teisė ir baudžiamasis procesas, kriminalistika, vadyba, įstatymų leidyba.

Vidmantas Egidijus Kurapka, Mykolas Romeris university, Faculty of Law, Department of Criminal procedure, professor. Research interests: criminal law and procedure, criminalistics, management, legislation.

Viktoras Justickis, Mykolo Romerio universiteto Socialinės politikos fakulteto Psichologijos katedros profesorius. Mokslinių tyrimų kryptys: kriminologija, įstatymų leidyba, psichologija.

Viktoras Justickis, Mykolas Romeris University, Faculty of Social Policy, Department of Psychology, professor. Research interests: criminology, legislation, psychology.