Data Loss Prevention in Higher Education

Petya Biolcheva[1], Miglena Molhova[2]

University of National and World Economy, Bulgaria[1, 2]

Abstract

The paper discusses data loss prevention (DLP) in higher education institutions, since data breaches are becoming more and more common in the recent years. Data compromised in higher education extend far beyond personal and financial data, including also sensitive research data. The environment of openness and collaboration between universities, as well as the typical access of many portable devices make access easier for hackers and detection of unauthorized access difficult. We argue that DLP is of high importance for universities, since they work with a variety of types of data, which are subject to different regulations and besides the necessity some of the data to be kept confidential there is also a counter need some of the data to be made available to the public (as results of academic research for example). The paper discusses how universities can define their critical data, the risks of data loss and strategies to keep their data safe. We discuss 3 main critical data loss prevention objectives, namely personal information protection/compliance, intellectual property (IP) protection and business partner compliance.The most common types of data breaches occurring in the higher education systems are hacking and malware, unintentional disclosure, and portable device breaches. The paper also presents a showcase of how Bulgarian universities address the issues described above.

Keywords: *data loss prevention, universities, risk management, intellectual property*

1. Importance of DLP for universities

In the recent years there has been a fundamental change in how university staff interacts with information technologies, powered by hardware (laptops and smart phones) and services (webmail, instant messaging, social media and remote access). Access to information remains the lifeblood of universities, but the risk of data loss is still significantly high. Data loss is a serious risk for all universities.[8] Universities are open organizations where people receive education, conduct research and exchange knowledge. At the same time, they process large amounts of sensitive data - personal data, student evaluation data, diploma data, contract information, projects, bank accounts, copyright, etc.[2] This information might be of interest to many individuals/organizations, some of which are deliberately working on malicious use of such information. The threat can come from both internal sources (abusive staff) and multiple external ones (hackers using various forms of phishing, viruses, unauthorized access to databases and servers, etc.). Data can flow from many different points. Loss, dissemination or modification of information can lead to asset damage, financial loss, outflow of students, loss of university reputation, or even bankruptcy. Demanding high security while maintaining educational services creates both challenges and new daily opportunities.[1]

In order to prevent data leakage and modification and to minimize the possibility of misuse, it is appropriate to use different combinations of rules, policies and tools that, when used correctly, significantly reduce the risk of data loss. In this article, we argue that this protection can be implemented through software solutions based on "Data loss Prevention" (DLP). At their core, DLPs are designed to counteract the risk of inappropriate disclosure of information. The main idea of DLP is to monitor and control the storage, movement or processing of data in general and critical and confidential data in particular, according to specific security policies. [7] Proper setup and implementation of a DLP system can perform functions of protection against data loss and unauthorized use of information. According to a research team that described the DLP application at its university the system helps analyze, monitor, track, lock and protect data. [2] The joint use of encryption and DLP protects a wide range of information: data about students, intellectual property, legal and financial records, correspondence with clients and partners, etc. The main position of the authors is that the application of a DLP systems in universities will significantly reduce the frequency of unintentional data leakage, will control vulnerabilities, information security will increase through the encryption capabilities of the system, which in turn will reduce external attempts for abuse.

2. Classification of data types used in Universities: defining critical data

Following the legal framework for the establishment and work of universities in Bulgaria [4], university information can be divided into 2 types: **Organizational information**, which is related to the organization, administration and management of the work of the educational institutions and the implementation of support activities. **Research information and information, related to the educational process -** it covers the scientific works, materials, research results developed by the academic staff, regardless of their form, for the servicing of the main activities of the educational institution. For the purposes of this paper, the scope of this type of information also includes scientific works, materials and research results developed by third parties outside the university, which are used in its teaching and research work, when these objects are subject to special regulation of the rights to use them and the university has obligations to protect the information in them (respectively to restrict access to them) at the request of the third party.

In order to implement DLP effectively, it is important for universities to be aware of exactly what information is critical for them and needs to be protected. According to the authors, the critical data of Universities need to include the following:

- Information related to the study materials, the assessment of the students, the diplomas of the students and doctoral students;
- Implementation of regulations of documents that must be protected from leakage, the internal rules of the university, including personal and sensitive data.
- Project documentation - projects on which scientists work at the university, which information is classified.
- Intellectual property rights - patents and related information on the application of inventions, rights to industrial designs, trademarks, know-how, etc.
- Unpublished and unprotected intellectual property materials and research results, the knowledge of which by a third party may compromise the research results or call into question their authorship or allow them to be used illegally, damaging the reputation of the university or blocking the possibility for further research and protection through the intellectual property system.
- Information that the university has acquired in the framework of an agreement with a third party (license agreement, non-disclosure agreement, etc.) and is obliged to limit its use to the scope and persons defined by the third party in the agreement.

3. DLP policies/strategies for Universities

When working on developing a DLP policy or strategy, universities must regard some main issues to start from:

- First they need to determine the primary data protection objective: this could be to comply with regulations (and often this is not something left to the choice of each organization, but is specifically envisaged in the legislation); protect the intellectual property of the university or comply with third parties. Of course depending on the needs of the university and the data it works with all these three objectives can be a part of the university DLP policy;
- Determine the main causes of data loss - one of the biggest challenges in mitigating data loss, is that there are so many reasons attributed to data loss in an organization and there is no tool or a simple solution that adequately address these various data losses. However to be able to address the risks faced, a solution must be developed to incorporate the causes of data loss, which are can be classified as people, processes and technology. [9]
    - People: Data loss can be caused by people through their lack of awareness of the security issues relating to sensitive information that are to be secured and most times are not been accountable for protecting these information.
    - Process: The process of securing these sensitive information can be caused by inadequate data usage policies, no proper data transmission process and lack of data monitoring usage.
    - Technology: Lack of flexibility and communication platform in technology deployed for the protection of data, makes it difficult for the user, thereby making the user to look for an alternative.
- Define the stages for developing a DLP system in the University – a basic DLP system consists of three stages which includes discover, monitor and protect.[6] The discovery stage locates where the critical data are been stored, by taking a detailed inventory of this data and then regrouping the sensitive data in terms of priorities. In the monitoring stage the organization should monitor how the confidential data are used, by understanding the content and context of this sensitive data and by analyzing when a breach occurs. The last stage which is the protect stage, basically

describes the ways for protecting data loss and this is done by being proactive in protecting these critical data.

- Define the DLP architecture – in general DLP architectures can be grouped in 4 main groups [3]:
  - endpoint DLP: Endpoint DLP relies primarily on purpose built software agents, that live on endpoints - laptops, desktops, servers, any device that runs on Microsoft Windows, Linux, or Apple OS X. The agent delivers visibility and, if desired, control over data. Deployment involves installing the agent on machines, where a protection is desired. No agent means no coverage.
  - network DLP: Network DLP, often referred to as agentless DLP, delivers visibility and control of traffic that passes across the network. A physical or virtual machine inspects all traffic, such as mail, web, IM and can then enforce data policies. Deployment is either via a physical appliance or a virtual machine then configuring network traffic to pass through for the inspection.
  - discovery DLP: Discovery DLP proactively scans your network, including laptops, servers, file shares, and databases to deliver a comprehensive analysis of where sensitive data resides on all these devices. To perform the data discovery some solutions require an agent to also be installed on the machine being scanned.
  - cloud DLP: Cloud DLP, much like Discovery DLP, scans storage repositories and delivers an accurate picture of where sensitive data lives, though as its name suggests Cloud DLP focuses on your data that lives in the cloud. Cloud DLP relies on an API (Application Program Interface) to connect with the cloud storage service (Box, One-Drive, etc.) then scan the content. Cloud DLP sees data as it is being put into the cloud and can perform a cloud storage audit or remediation.

When building a successful DLP strategy and the DLP systems related to it, universities must take into account that the DLP systems cannot function effectively in isolation. For a DLP system to effectively function it requires linking other security information processes as well. However, before implementing any DLP system, there is need to adequately understand what critical data the organization wants to hold, where is critical and confidential data being stored in terms of locations and the destination and the channels this information will pass through. So in order to make a DLP strategy successful universities should consider their own data life cycle. The data life cycle is a detailed outline of the phases involved in effectively preserving and managing of data to be used and reused. These stages include data at rest (data in storage), data in use (data flowing through internal and external networks) and data in transit (data that are been accessed).

4. DLP in Bulgarian Universities

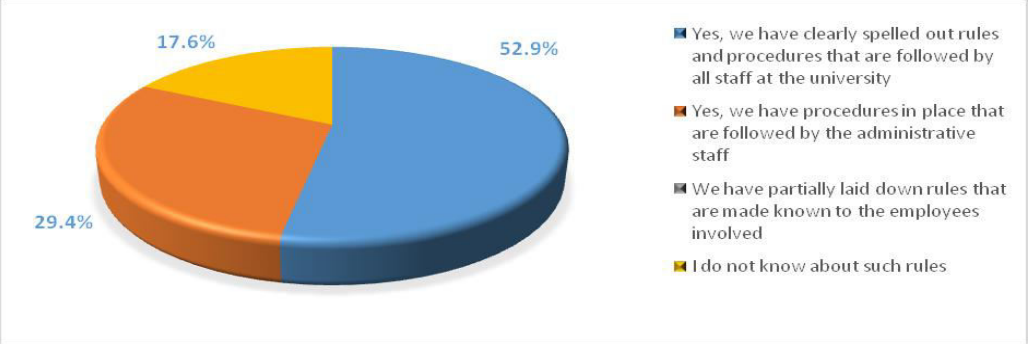
The methodology of the research is based on the conducted literature analysis and is organized in the form of a survey. The survey was held in 18 (35%) of the universities in Bulgaria. Respondents are professors, representatives of universities. The survey was organized through a structured online questionnaire, which provides anonymity of the answers, inviolability of personal opinion and ethics and relatively high certainty in the results obtained.[5] The study was conducted in the first half of 2020. It covers the manifestation of various risk situations related to information and data loss for a period of five years ago.

The main fields covered by the study are: the presence of negative experiences related to information and data loss; rules regarding the work with different categories of information, ensuring the protection of information, protection of intellectual property at the university. The methods of descriptive statistics were used for the analysis of the data.

Regarding the problems related to information leakage in the surveyed universities, three main types of sensitive information were identified. The first is related to the data related to the assessment of students and the possibility for the assessments to be manipulated. 35.3% of the respondents have faced this risk in the last five years. Next, the study analyzes the data loss related to public procurement, research projects, test materials and more. Although with a smaller percentage (about 12%), but again respondents confirm there is information leakage.Another problem is the forgery of university diplomas - aalthough the percentage is relatively small at around 6%, this risk remains a major threat for the reputation of universities.

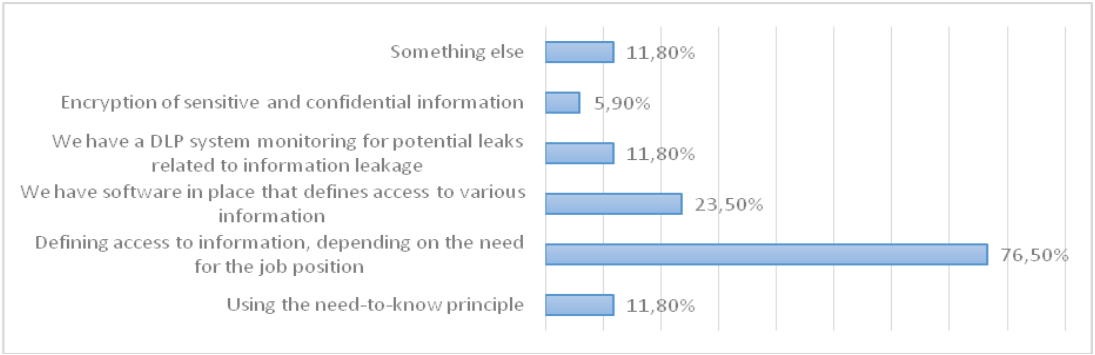
The next study area refers to the work with different categories of information, rules and procedures for processing information, its storage and extraction. Half of the respondents indicate that their universities work with clearly defined rules and procedures, through which the different categories of staff have access to the information necessary for the performance of their work duties. They follow established mechanisms for access, processing, destruction of information. They declare low levels of

data leakage. In another 30%, the procedures for working with different categories of information relate mainly to administrative staff, with access to sensitive and classified information. The third part of the answers (about 20%) is worrying, in which the respondents are not aware of the existence of current rules and procedures in this field. This means that they are missing or are not available at all. This fact significantly increases the potential risk associated with data leakage from universities (see Graph 1).



Graph 1 Existence of rules and procedures for working with different categories of data in Bulgarian Universities

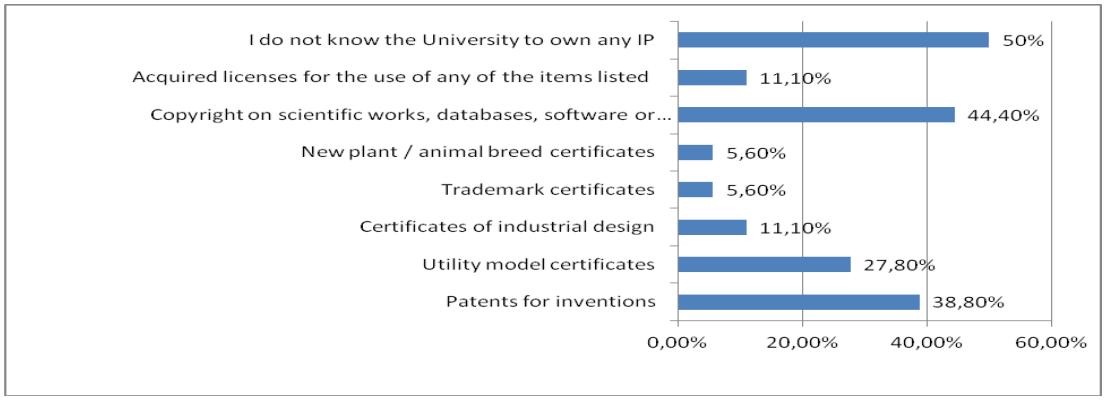
To determine the extent to which these rules and procedures were actually followed, respondents had to answer a question concerning information classified as sensitive to which they all had access. These are the data related to the evaluation of students and the channels of dissemination of this information. For about 6% of respondents, the possibility of information leakage is limited due to the fact that the exams are conducted electronically, the grades are generated automatically, and the access of the teachers is limited. In almost 30% of the universities surveyed, teachers have access to exam grades and enter them into specific software themselves. Here the potential risk of data leakage or modification is greater, but the protection that is offered is good. The traditional way of entering grades in student books by the teacher is used by about 40% of the respondents. This method reveals various weaknesses in completing, accepting, submitting evaluation protocols. Worrying is the fact that there are still universities where there are cases of sending group emails with names, faculty numbers, and other personal data to students. This type of data dissemination violates the provisions of the GDPR and is in itself a leakage of personal data.
All surveyed universities apply measures to prevent information leakage. Most respondents work with clearly defined rights regarding access to different categories of data. About 35% of them have security software that reduces the possibility of information leakage. Among them (about 12%) rely on the DLP system offered in the present material (see Graph2).



Graph 2 Prevention of information leakage from universities

When it comes to the intellectual property the universities own we can see that 50% of the respondents say that they do not know if the university owns any IP. The people, who know about the IP of the university share that the greatest percentage of the IP is vested in copyrights (44,4%), followed by patents for inventions (38,8%) and utility model certificates (27,8%). Of course when analyzing the results we must take into consideration the type of scientific work the universities implement, since not all scientific work is related to research and development activities, leading to innovations, which can be protected as inventions, industrial designs, utility models. The more

surprising result is that do not recognize the importance of trademark rights as an instrument to be used on the educational market –for being recognized and for building reputation.(See Graph 3)



Graph 3 Intellectual property in Bulgarian Universities

Related to the IP rights the university owns we wanted to know if the universities have any IP protection policies and measures. We asked employees if they are aware of actions they can and cannot perform in relation to the protected IP objects – if they have been instructed, given rules to follow. But only 50% of the employees answered yes, so Universities still need to work hard in that domain to clearly address IP issues to their employees in order to be able to more effectively protect it. The graphs below show what knowledge employees have in terms of working with the university IP.
As it can be seen from the graphs only a small percentage (around 30%) of employees know about procedures they need to follow in terms of what to do if they create an IP work/object or to prevent outsiders from infringement of the university's IP rights. Universities must seriously consider this field of DLP, since most of their work is anyway locked in intangibles and if there are no proper measures to implement an IP policy even if there is one, then the risks of breaches and IP infringement increase immensely. Universities also work with a lot documents with restricted access - 61% of respondents say that they work with such types of documents in their organizations. However almost half of the respondents (44,4%) say that they can freely transfer information from the university devices they work on to an external device. 28% of the respondents answered that they need a specific permission to do that. The next graph shows what a high percentage of the employees do not know if the university has a system to notify them if there is an attempt by an outsider/outside organization to access their documents or research (over 50%). Yet 61% of employees say that they know who to contact in case of such event. 66,7% of employees consider that it is appropriate for the university to have a DLP system in place to monitor indications of information abuse.

References

[1] Adeniyi, W., Fatokun, J., The Risk of Data Leakages as Data Grows in Information Technology-Driven Tertiary Education Institutions, European Journal of Mathematics and Computer Science Vol. 5 No. 2, 2018, p. 33
[2] Boranbayev, A., Mazhitov, M., Kakhanov, Z., „Implementation of Security Systems for Prevention of Loss of Information at Organizations of Higher Education", 12th International Conference on Information Technology - New Generations, 2015
[3] Digital Guardian, "The definitive guide to data loss prevention", 2017
[4] Higher Education Act, Promulgated, State Gazette No. 112/27.12.1995, last amended No17/25.02.2020
[5] Jamsen, J., E-Survey Methodology, from Handbook of Research on Electronic Surveys and Measurements, Chapter 1, 2017
[6] Jonathan Jesse and ITS Partners, "Symantec DLP Overview", Available at: http://www.symantec.com/en/uk/business/theme.jsp?th, 2015
[7] Salem, M. B., Hershkop, S., and Stolfo, S. J., "A survey of insider attack detection research," in Insider Attack and Cyber Security, vol. 39 of Advances in Information Security, Springer US, 2008, pp. 69–90,
[8] Stringer, J.,"How to implement a data loss prevention strategy", Sophos, 2010, p.1
[9] Waziri, et al., „Data Loss Prevention and Challenges Faced in their Deployments", International Conference on Information and Communication Technology and Its Applications, (ICTA 2016)