Synthesis
International Scientific Conference of IT and Business-Related Research

PRIVACY POLICY AND DATA ARCHIVING IN ORGANIZATIONS IN THE REPUBLIC OF SERBIA AND THE EU COUNTRIES

ZAŠTITA PRIVATNOSTI I ARHIVIRANJE PODATAKA U ORGANIZACIJAMA U SRBIJI I ZEMLJAMA EVROPSKE UNIJE

Dušan Stamenković, Marko Šarac, Dalibor Radovanović, Ana Simićević
Singidunum University, Belgrade, Serbia

Abstract:
The way in which an organization protects and archives business data often causes great controversy, without clearly established rules on the level of countries and organizations. The issue of privacy violation, and the outflow of confidential information about the citizens employed at governmental institutions that should provide the highest level of protection. Such developments are not new, as user rights injuries are the phenomena occurring not only in Serbia, but also globally. In recent years, we have been witnessing numerous scandals involving the world's largest companies such as Apple, Dropbox, and Google. The laws of the Republic of Serbia prescribe to some extent how to perform backup and data protection procedures. The authors have dealt with the issue of data storage and enabling of backups in large organizations without compromising data integrity.

Key words:
data protection, backup, archiving, RAID, DPM.

Apstrakt:
Način na koji organizacija štiti i arhivira podatke često izaziva veliku polemiku, s obzirom da ne postoje jasno definisana pravila na nivou država i organizacija. Sve češći su slučajevi narušavanja privatnosti, kao i odliv poverljivih informacija o građanima zaposlenim u državnim institucijama koje bi trebalo da obezbede zaštitu na najvišem mogućem nivou. Ovakva kršenja prava korisnika nisu novina samo u Srbiji, već su zastupljene na globalnom nivou. U proteklih nekoliko godina, svedoci smo velikog broja skandala vezanih za najveće svetske kompanije kao što su Apple, Dropbox, Google. Zakoni Republike Srbije u određenoj meri propisuju procedure koje se tiču zaštite i čuvanja (back up) podataka na računaru. Autori se bave pitanjem skladištenja podataka i čuvanja podataka u velikim organizacijama bez ugrožavanja integriteta podataka.

Ključne reči:
zaštita podataka, čuvanje podataka (back up), arhiviranje, RAID, DPM.

1. INTRODUCTION

A new law on data protection in the EU (European Data Protection Act, 2012) was enacted in 2012, and started with implementation in 2014. It represents an upgrade of the 1995 law. Its essence is to bring together some of the different laws in different countries. Some of the key provisions of this law are:

- It is necessary that all companies and organizations with more than 250 employees have the person in charge of data protection (Data Protection Officer)
- Companies are required to submit an issue and notice, within 24 hours following data leakage
- If personal data (Personally Identifiable Information - PII), referring to the EU's population, are sent outside the EU, the local agency (body) data protection (Data Protection Authority - DPA) must be informed. In the case of the Republic of Serbia, the institution of the Commissioner for Information of Public Importance and Personal Data Protection.

Also, companies must pay particular attention to:

- Planning, standardization and compliance with the rules relating to maintaining the data in the design, construction and development of business systems,
- Operating systems should be designed so as to provide the highest level of security and preservation of data and making data backups,
- Once a year organizations have to undertake a detailed assessment of the risks of data loss and propose steps to reduce such risk.

There is a number of laws of the Republic of Serbia related to the protection and preservation of data, which stipulate that certain documents must be kept for a statutory period of time. The most sensitive data are mostly related to the documentation related to the company's operations and finances. Retention periods are shown in Table 1.

Type of documentation	Retention period
Financial statements (balance) and reports on audit	20 years
General ledger	10 years
Sub ledgers	5 years
Payrolls and analytical records of salaries if they represent important data	permanently
The documents on which the data are entered into the books	5 years
Sales and control blocks, forms and similar documents	2 years
Tax (VAT) records	10 years

Table 1. The time period of keeping documents

A large number of documents are stored in a physical form (manuscript, printed record, diary, *etc.*) and the proposal of

E-mail: dstamenkovic@singidunum.ac.rs
DOI: 10.15308/Synthesis-2015-71-75

a new law of archives and archival services of the Republic of Serbia passed a series of polemics. First of all, the opinion of economists is that the application of the law can block the development of electronic business and commerce, and that it is necessary to adopt a special law on the preservation of electronic documents. The Law on Electronic Documentation and Electronic Signature Act passed in 2009 and 2004, as well as related by-laws regulate only issues of production, receipt and delivery of electronic documents, and not the question of permanent storage and transfer to the competent institution preservation, *i.e.* archives. In this way, companies are doomed to independently implement policies for archiving electronic data within their organizations.

2. PROBLEMS AND CURRENT SOLUTIONS

The issue of privacy policy and data archiving in organizations is not new and other authors have dealt with it.

In one of his papers, the author Emsley, Rob introduces Policy-Based File System Archiving. Escalating growth of file systems underscore the need for a practical solution that embraces and enables data growth while optimizing storage costs and meeting retention and access requirements. This can be accomplished through the use of tiered storage architecture. Tiered storage places critical, timely files on high performance redundant array inexpensive disks (RAID) or network attached storage while access needs are high, and then seamlessly moves them to secondary storage when access needs decrease. A tiered storage infrastructure is made up of multiple types of disk tape and optical storage. Policy-based file system archiving matches the right data to the right storage at the right time using customer-driven business rules. This is a core component of an information life cycle management strategy. With policy-based file system archiving, administrators can define rules to identify files to be moved to a more appropriate level of storage. File candidates might include data that is inactive or that needs to be archived for compliance or governance requirements. Additionally, policies can be established to move files between storage tiers over the lifetime of the data. Customers can precisely target files better suited to reside on alternate storage tiers as well as manage data placement across storage tiers to match service levels. As a result, organizations can effectively manage data movement, access requirements, storage costs and capacity needs (Emsley, 2005).

In another paper, the author Ginty, Ed, discusses which media is best for archiving and backup policy. The author describes the traditional methods used by many organizations in data archiving which include the WORM technology, disk-to-disk transfer, as well as the use of back-up tapes and purpose-built, secure archive devices. According to the author, the use of the proper data-archive technology is important because it can provide substantial savings to organizations, not only in terms of money, but also in terms of time, operations, business disruption savings, as well as emotional cost savings (Ginty, 2007).

Most actual paper on the topic of archiving and backup titled "Big data, open government and e-government: Issues, policies and recommendations" is written by authors Bertot, John Carlo, Gorham, Ursula, Jaeger, Paul T., Sarin, Lindsay C., Choi, Heeyoon. The transformative promises and potential of Big and Open Data are substantial for e-government services, openness and transparency, governments, and the interaction between governments, citizens, and the business sector. From "smart" government to transformational government, Big and Open Data can foster collaboration; create real-time solutions to challenges in agriculture, health, transportation, and more; promote greater openness; and usher in a new era of policy- and decision-making. However, there is a range of policy challenges to address regarding Big and Open Data, including access and dissemination; digital asset management, archiving and preservation; privacy and security. After presenting a discussion of the open data policies that serve as a foundation for Big Data initiatives, this paper examines the ways in which the current information policy framework fails to address a number of these policy challenges. It offers recommendations intended to serve as a starting point for the revised policy framework to address significant issues raised by the U.S. government's engagement in Big Data effort (Bertot *et al*., 2014).

BACKUP AND ARCHIVING METHODS IN THE REPUBLIC OF SERBIA

Electronic document is a set of data consisting of letters, numbers, symbols, graphics, audio and video files contained in the application, a written document, document or any other document drawn up by legal and physical persons or authorities for use in legal transactions or administrative, judicial or every other proceedings by the authorities, if made electronically, digitized, sent, received, stored or archived in electronic, magnetic, optical or other media (Draft Law of Archives and the Archives Service of the Republic of Serbia, 2012). The main elements of the documents are authentic, authenticity, integrity and usability, in accordance with the definitions provided within the international standard ISO15489 for document management.

Small and medium-sized enterprises often encounter problems with archiving of electronic documents, or the lack of a capable professional staff in the field of information technology. Some of the methods of making backup / archival copies of data are copied (backup) to external memory (USB Flash) or storing data on CD/DVD.

The practice in large enterprises is to primarily follow the procedures and standardized processes while protecting and archiving data in order to meet the main elements of the document. As regards this, large companies resort to technical and technological systems and different document management solutions that comply with the international standard ISO15489 SRPS.

The basic methods used as protection against data loss are:

- Use of virtualized data warehouse,
- Automated backup to a local server,
- Archiving data to a remote server (Disaster Recovery)
- Use of cloud storage systems.

Companies with IT support usually use advanced technology to protect data integrity and smooth continuation of business due to cancellation of part of computer equipment. The basic method for protecting data integrity is RAID (Redundant Array of Independent Disks) technology. Redundant Array of Independent Disks is now used as a term for data storage due to which data can be shared or repeated between two or several physical disks. There is a number of different levels of using this technology. Some of these levels are being used by the workstations, and some are server-side. The basic idea of RAID1 level is to increase reliability and enable replication of data from one disk to another in the array. In practice, this technology is in its original form and it allows you to lose a single physical disk in an array without any consequences in the form of data loss. If one drive fails, all the lost data can be recovered from the second disk in the array. The use of RAID1 technology is shown in Figure 1.

DOI: 10.15308/Synthesis-2015-71-75

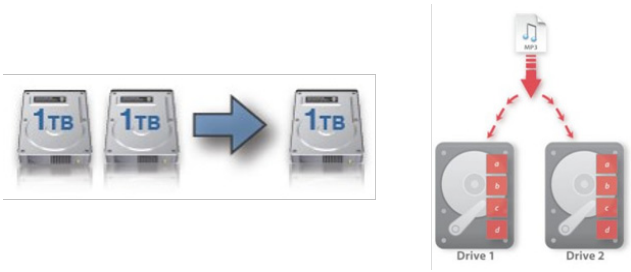

Figure 1. Schematic representation of RAID1 technology

The possible problem for configuration of workstations like this is that they require greater financial investment in the form of procurement of computers that support this technology, but also twice the number of disks in relation to the size of the data storage (*e.g.* Two drives of 1TB per allow 1TB of storage). The most important thing to consider when introducing this technology is that it does not constitute archiving or backup of the data.

Newer operating systems that are mainly used in the business environment such as Microsoft Windows (versions 7 and 8) provide an opportunity to the additional options. Shadow Copy and File History enable the automation of creating data backups. These two technologies, depending on the operating system, create copies of documents on a particular system partition hard drive or on another physical hard drive specified for backup.

It takes time to create these copies, which can last from several minutes to several days depending on the effects that the user wants to achieve. It is important for the understanding of this type of backup process that the data is not copied again and again. The copy of files is not kept countless times, they are copied only once, and each subsequent time the difference between the original document and the document is copied. This kind of copying and archiving of files is called incremental copy and it enables rational use of storage space. Storage space reserved for the backup may vary. Depending on the determined capacity of the hard drive, the size and the number of documents and the time interval depends on the number of hours, days, months and even years the user can go back in time and look at the different revisions of backup document. Using Shadow Copy and File History, display of the backup data at different time intervals can be viewed in Figure 2.

As already mentioned, safety ensured copy (backup) can be taken manually, and it can be configured as an automated backup process on each workstation individually with the help of Shadow Copy or File History option. In large enterprises with 250 or more employees, this type of backup processing of data represents a waste of time and resources. These organizations resort to the implementation and use of automated technological systems and solutions for data protection and archiving. Some of the Microsoft Windows Server centralized solutions that implement policies of automated backup data process are;

- Roaming profiles on centralized Active Directory,
- System Center Data Protection (Data Protection Manager - DPM)
- Products by third parties such as Acronis, Yosemite, *etc.*

This paper describes the operation of the System Center Data Protection Manager (DPM) on Windows Server 2012 R2 Active Directory domain. System Center DPM is a product of Microsoft Corporation that provides a reliable method for data recovery, backup and near-continuous data protection in a Microsoft Windows environment (Microsoft System Center Data Protection Manager Technical Preview, 2005). DPM also uses Shadow Copy technology for continuous backups. If DPM uses the same method for backup, and obviously does the same thing as a backup on single Windows OS, do we really need DPM?

The main purpose of DPM is that it delivers centralized backup of a large number of computers in an organization. DPM also provides the following services:

- More efficient data protection
  - DPM capture all changes as they happen instead of copying and replacing the entire files. DPM backup does not need huge load on production servers unlike conventional backup tools. DPM doesn't have to copy entire files if even a single byte changes, as in that case, only changed bytes will be copied.

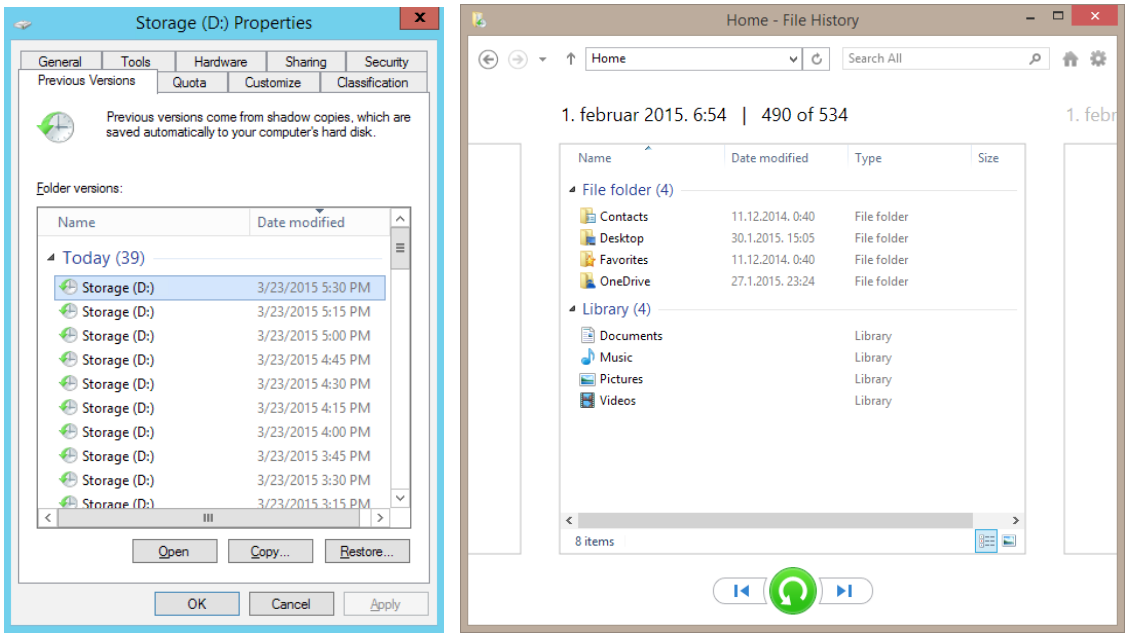

Figure 2. Comparative review of backup data using Shadow Copy and File History option

DOI: 10.15308/Synthesis-2015-71-75

- Flexible backup scheduling
  - Backup frequency can meet the needs of an organization; it can be hourly, daily or weekly. IT administrators can specify data sources and the type of data or file that are suitable for backup.
- Multiple point-in-time backups using frequent shadow copies
  - Shadow Copies are also known as snapshot, point-in-time, past-time, or cache copies. Using Shadows Copies, it allows IT administrators to browse and recover the deleted or corrupted files from the period designated for storing copies of documents. This period may vary from few hours to few years; the limiting factor is the size of hard disks for storage backups.
- Protection from network outages and hardware failures
  - The DPM agent installed on computers, storage logs and backups of the data from each of the computers to the server running DPM. If the computer goes down or it is destroyed, a copy of the data is still available on the server running DPM. If the network goes down, the agent that is installed on the computer continues to cache all the changes until the network is working again.
- Network throttling
  - DPM uses its own metrics control to avoid throttle network traffic, in this way, DPM can minimize impact of copying data across network.
- Integration without disruption
  - DPM uses the existing backup systems to increase efficiency of the data protection infrastructure and lowers the total protection costs.

RAID disk technology offers significantly better I/O performance, combined with the steadily increasing mean time between failures (MBTF) of hard disks and controllers, using DPM to leverage RAID storage as a backup mechanism yields superior protection and speed.

- Rapid and reliable data recovery from disk
  - Because DPM is disk-based, it can take advantage of the additional reliability benefits of a RAID for protection.
- Automatic error correction
  - DPM monitors its backup jobs. This process is needed to ensure that jobs are completed without error. If any error is detected, two-stage process for error correction will begin. The first part is that the replica on DPM server automatically validates against the data on computer to ensure that the replication is consistent and has occurred as planned. The second part of the process is to check inconsistencies between the data source and its replica. If the process finds errors during validation, the fix-up activity re-sends the object(s) from the data source (computer) to the replica (DPM server).
- Recovery options
  - The most common scenario in an organization is the request sent by the end-user for recovering lost, deleted or corrupted files to IT technical support to find a backup or archived version of the files. Assuming that the business has an archiving system in place, DPM provides practical recovery alternatives, including the ability for administrators to delegate recovery to end users and the ability to recover the contents for the user individual files and folders.

Once DPM is deployed on the dedicated server, *i.e.* on every computer in domain network which needs to be protected, the DPM agent must be installed on it (Fig. 3). After remotely install of agent, computer is still not protected. Administrators must create protection group and define policy concerning the time and frequency of backup on a daily basis, as well as to define how many days backup will be kept on the server and how much storage space every protected computer can assume. When all the conditions are met, the computer that needs protection can be added to the group, after which synchronization and data verification on that computer will begin (Fig. 4). After some time, depending on the number of files and folders the protected computer has, backup files will be available through recovery options (Fig. 5).

3. CONCLUSION

Data protection and data archiving are not always the primary fields of interest for an organization, which is particularly

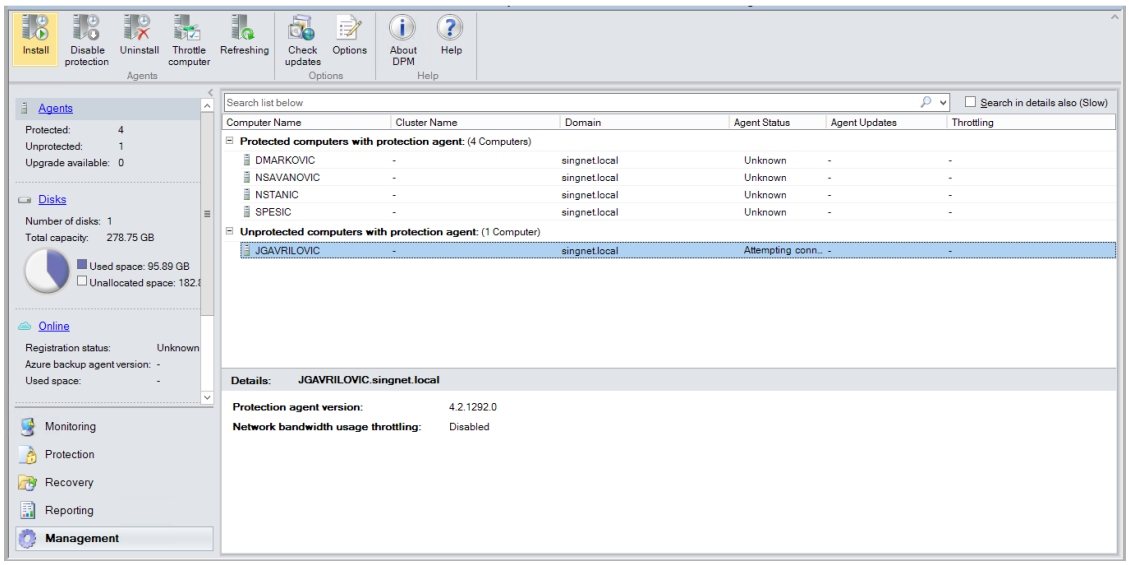

Figure 3. DPM shell with the installed agent list

DOI: 10.15308/Synthesis-2015-71-75

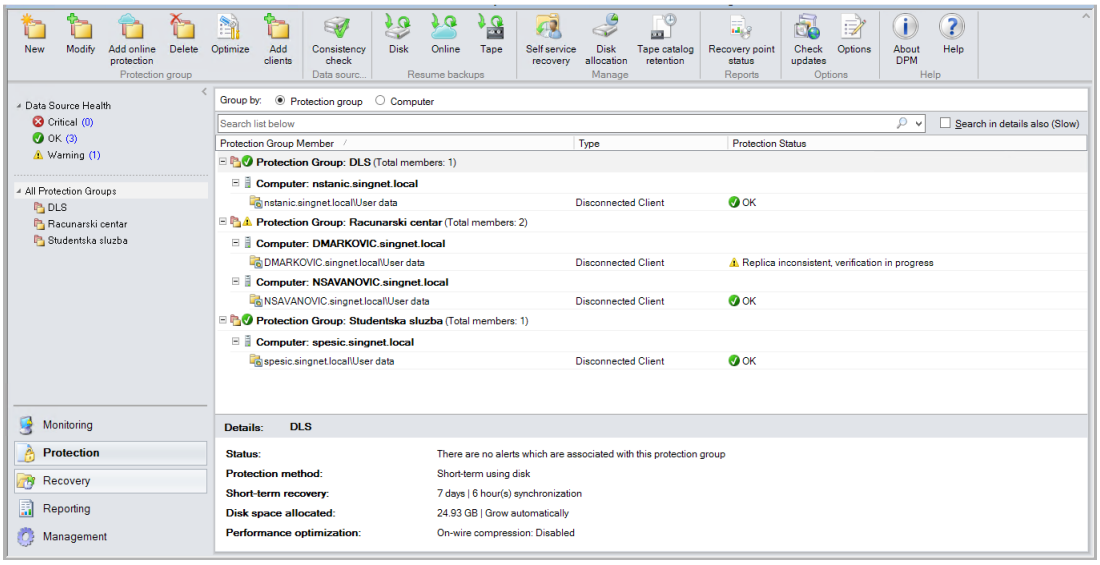

Figure 4. Protection groups and protected computers status

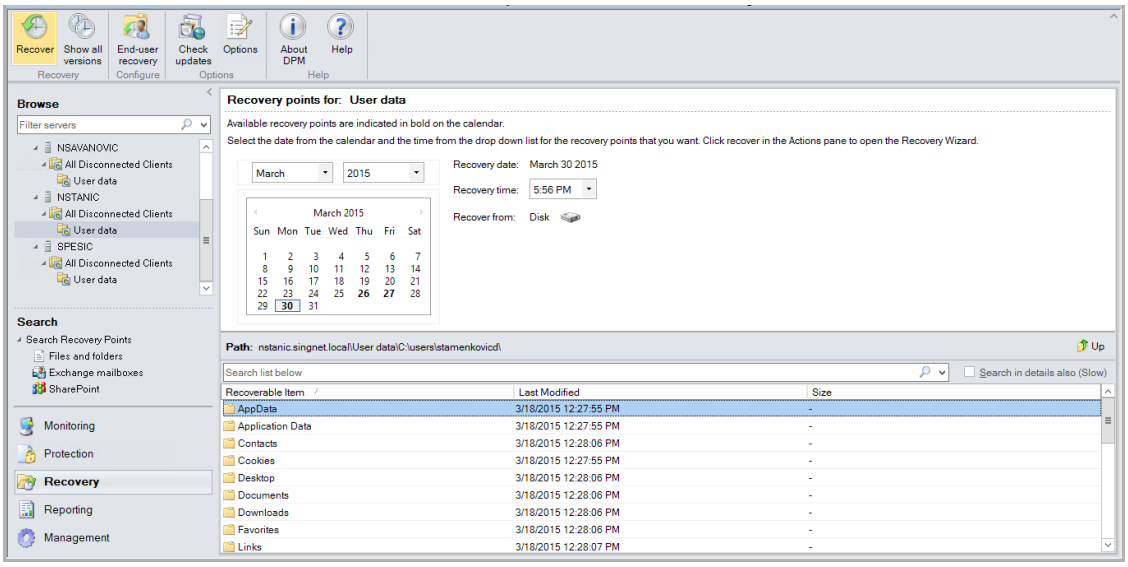

Figure 5. Recovery points of backup files

the case in the Republic of Serbia. As regards the data integrity, data loss or data leakage becomes a matter of concern. This is the situation in which no organization wants to be. However, when it occurs, data protection and data archiving policy and practice are considered useful and preferred by an organization. The authors have come to the following conclusions:

- Hardware RAID 1 implementations offer real-time physical hardware failure protection at the local and server level.
- Shadow copy and history functions offer real-time and periodical backup and archiving functions at the local and server level.
- DPM offers low ongoing administration costs compared to conventional backup systems because even end-users can recover their lost files, unlike before when the job was done by administrators.

After rigorous testing and practical implementation of multiple systems and practical work at the University IT center, the authors reached the conclusion that Microsoft DPM offers multiple ways of data protection based on real-time and periodical backup and archiving functions at the server level. When combined with RAID 1, DPM represents the recommended solution for an organization's data protection policy.

REFERENCES

Bertot, J.C., Gorham, U., Jaeger, P.T., Sarin, L.C., & Choi, H.Y. (2014). Information Policy & Access Center. *The International Journal of Government & Democracy in the Information Age, 19*, 5-16.

Narodna skupština Republike Srbije. (2012). Zakon o arhivskoj građi i arhivskoj službi. Retrieved January 05, 2015, from http://www.parlament.gov.rs/upload/archive/files/lat/pdf/predlozi_zakona/3184-12Lat.pdf (In Serbian).

Emsley, R. (2005). EMCs Information Management. *DM Review, 15*(9), 38-39.

European Commission. (2012). European Data Protection Act. Retrieved January 05, 2015, from http://ec.europa.eu/justice/data-protection/

Ginty, E. (2007). Secure Data Group. *Information Systems Security, 16*(2), 90-92.

Microsoft. (2005). System Center Data Protection Manager Technical Preview. Retrieved January 05, 2015, from http://www.microsoft.com/en-us/download/details.aspx?id=44304

DOI: 10.15308/Synthesis-2015-71-75