TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

ACTA TECHNICA NAPOCENSIS

Series: Applied Mathematics, Mechanics, and Engineering
Vol. 64, Issue Special IV, December, 2021

QUALITY AND DIGITALIZATION OF THE DATA BACKUP PROCESS MANAGEMENT IN ORGANIZATIONS HAVING AS OBJECT OF ACTIVITY THE PROTECTION OF INTELLECTUAL PROPERTY

Mihail Aurel TITU, Radu Costin MOISESCU, Constantin OPREAN

Abstract: *The proposed paper argues from a scientific point of view the need to implement centralized data storage technologies on disks and digitize related processes in intellectual property organizations. The paper addresses the issue of providing centralized data storage platforms in the context of the need to integrate all physical or virtual information systems of these organizations. The research refers to an analysis of both the processes of centralized saving of this data on disks and the database security technologies that may be applicable in the information systems within these organizations. The issue of analysis of architectures and technologies that can be applied in the field of securing and saving databases in existing information systems in these organizations but especially improving management and increasing the quality of execution of these processes are extremely relevant aspects analysed and explained in this research. The proposed research can be considered a reasoned and documented point of view in a topical field.*

Key words: *information system, computer system, data architectures, centralized backup, disk backup*

1. INTRODUCTION

In organizations having the object of activity in the protection of intellectual property, disaster backup and recovery processes should be implemented at a level that ensures that they are sufficient and implementable in the event of a data and database protection incident. date. But recent studies (2016 study on the cost of data breach by the Ponemon Institute) indicate that, in general, these organizations do not apply sufficient measures to ensure that data backup and recovery procedures are implemented and tested rigorously, the management's attitude being based more on the principle that "this will not happen to us".

The main organizational and security factors and characteristics that may affect the data of these organizations are:

a. Cybercrime representing the result of a malicious or criminal attack;
b. Threats within the organization - data breaches caused by intentional encryption, attacks, deletion, corruption and theft of information by its own employees;
c. Hardware failures caused by outdated or poorly managed computer systems;
d. File corruption and backup and recovery software failure;
e. Human error - studies indicate that 27% of data loss is due to human error;
f. Natural disasters and extreme weather.

In this context, the implementation of adequate cyber security measures also significantly reduces the average cost of possible data loss or corruption, thus leading to a significant improvement in the management and quality of centralized backup processes [1]. One of the measures that can be implemented to reduce the risk of data loss or corruption in these organizations is to integrate a centralized data backup system either on disk or on magnetic tape type LTO6 [2].

Thus, the data backup solution that is proposed to be implemented through the continuity plan in the infrastructure of an intellectual property organization will be used to save the existing virtual environment in the infrastructure as well as filesystem data and specific Informix integrations.

2. DATA BACKUP PROCESSES. CONTRIBUTIONS TO IMPROVING PERFORMANCE.

The centralized data backup solution that is proposed for implementation in an intellectual property organization, provides a centralized backup platform for all IT systems in the card of the institution, physical and virtual, as well as advanced integrations with hardware equipment in the category of centralized storage (Storage Area Network), backup on disk or magnetic tape systems LTO6 [3]. The solution uses an LTO6 magnetic tape library for backup image storage, as well as licensed disk space for in-application use within the TB limit useful before compression and deduplication on Data Domain DD6300 equipment configured as a Data Domain Boost destination [4].

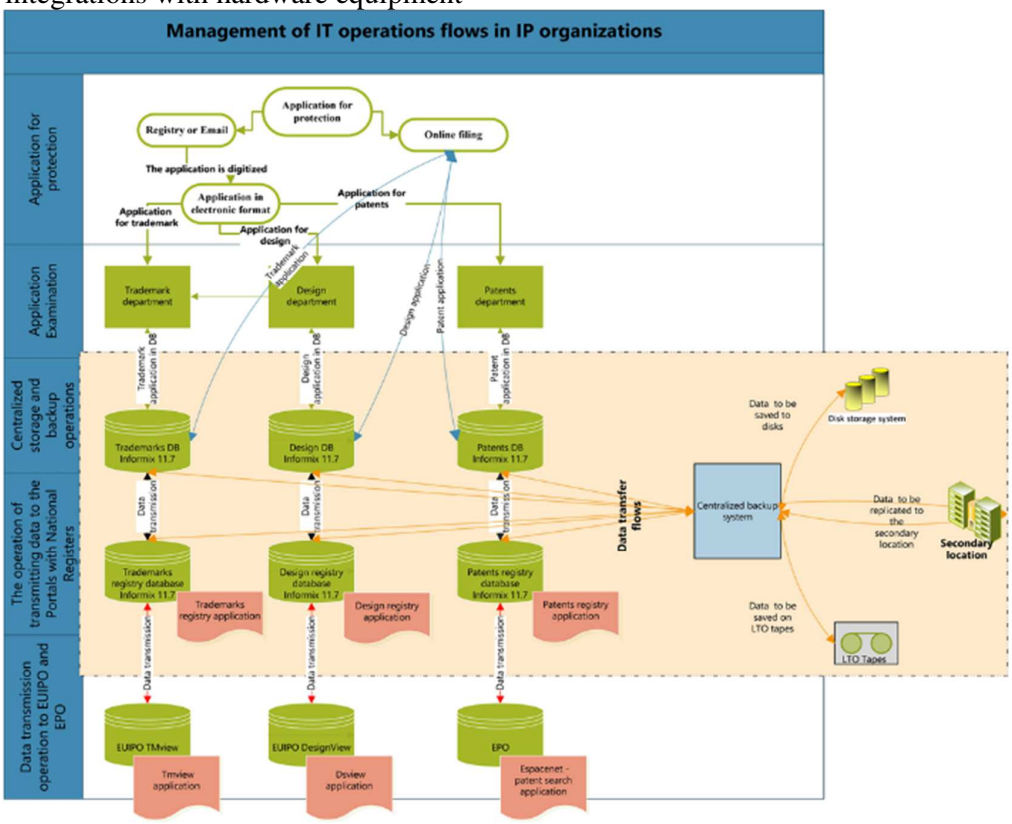

Fig.1. The main data flows in IP organizations

The centralized backup application will install agents for specific operating systems and applications with which it will integrate and which will be described below. The application also manages LTO6 cassette backup media, both those in the library and those taken out in off-line storage procedures and stored outside the library.

Figure 1 shows the main data flows existing in these types of organizations. This is how we define the following data streams starting with the submission of applications for protection titles:

a. Digitization of applications for situations when they are submitted on paper or sent by mail or email. The digitization process includes scanning and/or entering applications into the institution's electronic system.
b. Applications in electronic format are entered in the databases specific to each category of protection requested: protection for trademarks, inventions or industrial designs.
c. Applications in electronic format are entered in databases specific to national registry applications for publication in order to offer the possibility third parties to formulate appeals - the flow of data for transmission to the portal that manages the national registers.

d. Applications in electronic format entered in the databases specific to each category of protection requested: protection for trademarks, inventions or industrial designs are examined in advance to verify whether they meet the minimum legal requirements (correctness of completion of forms, payment of fees, etc.).
e. Applications are examined in the second phase to determine whether they meet the legal conditions for the granting of protection titles.
f. All data on the preliminary and second phase examination stages as well as the granting of protection titles form the data flow through which the international databases are daily updated. These databases are managed by European Union bodies and agencies such as EUIPO (European Union Intellectual Property Office) for trademarks and industrial designs, EPO (European Patent Office) for patents and worldwide organizations as WIPO (World Intellectual Property Organization).
g. The data flow that is transmitted to the proposed centralized backup system, namely backup to disks, on LTO6 type magnetic tapes and the one transmitted through the replication mechanism to a possible secondary location.

The purpose of implementing a centralized backup system is to improve the main parameters specific to these processes: reducing the time to complete backup sessions, reducing the share of failed sessions and reducing the time required for data restoration after a disaster [5]. Thus, the increase of quality and the development of a management based on monitoring and detailed reporting of the execution of these processes are the result of improving the parameters related to the main backup and restoration processes with immediate effect [6]. Within the centralized backup system, we define the management, support and main processes presented in a global view in Figure 2.

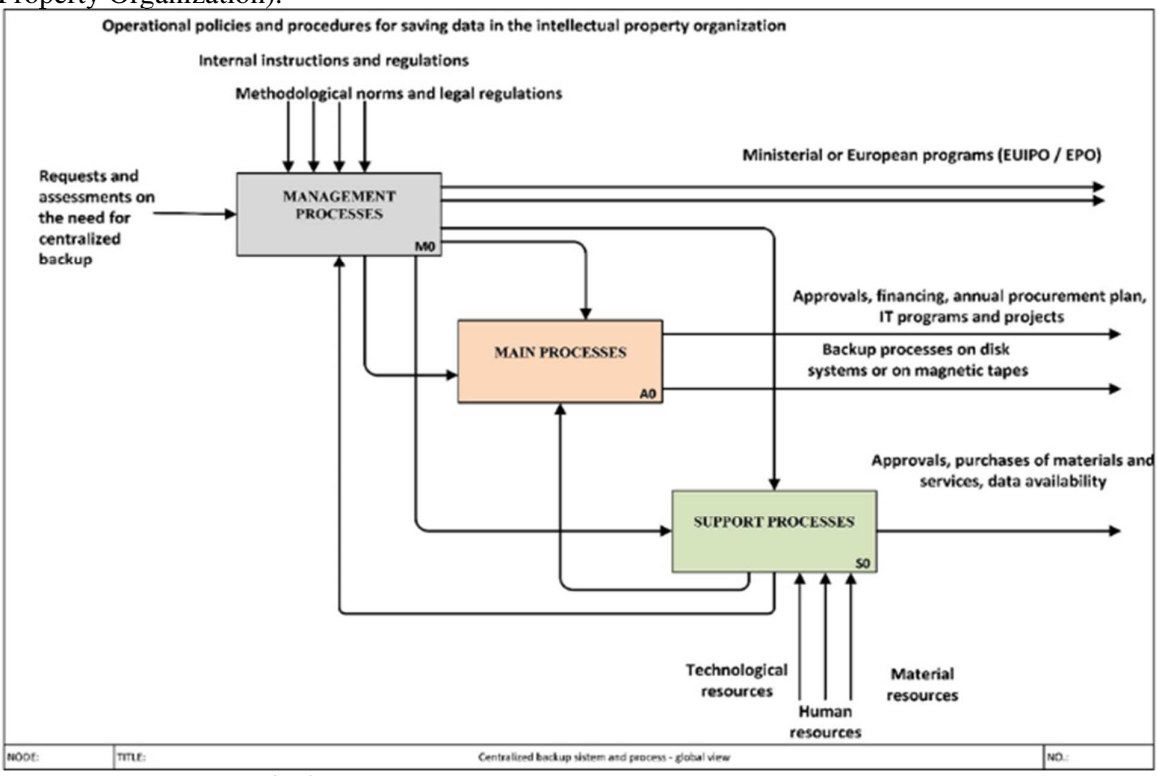

Fig. 2. The main processes within the centralized backup system

Thus, when a data backup session is started, a record is created in the internal database. of the backup system, which depending on the session configuration (full or incremental) will send the data to be saved either on the LTO6 magnetic tape system or on the SSD or SAS disk volumes

system [7]. Also, for each object in the session, an object version record is created. Both records are stored in the CDB (catalog database) part with multiple attributes. Backup operations will be scheduled to run automatically and will be permanent monitored and verified by a backup administrator. The correct and error-free execution of the backup sessions will be checked to take any remedial action, if necessary, with the possibility of adding any other backup jobs in the future as needed and in compliance with licensing and space limitations for backup [8].

All media records are stored in the Media Management part and are allocated for backup according to the defined backup policies. When a data segment (and a catalog segment after it) is written to tape, a media position is recorded in the CDB for each version of the object that was part of that data segment. In addition, the catalog is stored in the Detail Catalog (DC) binary file. The catalog name is not changed when save sessions are attached to the same environment. If a medium is overwritten during a save session, its old DC binary is deleted and a new DC binary is created [9]. In order to implement the backup policy, it is proposed to implement a scheduling scheme that allows saving all data and databases and which is described below:

a. on Monday-Thursday, an incremental Daily Backup is performed for the filesystem, IDB, Infomix and Hyper-V virtual environment specifications [10].
b. Every Friday, a Weekly / Monthly / Yearly Backup Full backup is performed for the filesystem, Informix and Hyper-V virtual environment specifications.
c. on Saturday-Sunday, all Weekly / Monthly / Yearly saves are copied on LTO6 type tapes.

All session messages generated during the backup process are stored in session message binaries. Backing up data is a process that creates a copy of system data on backup media. This copy is stored for later use if the original data is destroyed or corrupted [11]. A backup session is based on the backup specifications and can be started interactively. During backup sessions the data must be consistent [12]. This data that needs to be saved is read, transferred over the network and written on various media (disks, magnetic tapes, complex systems in other locations). Among the capabilities of the data backup system proposed to be implemented we highlight:

a. Automatic balancing of device use (load balancing);
b. Automatic saving of shared disks;
c. Combining “full” type backup sessions with incremental ones in order to save time and media support;
d. Allowing the organization of backup copies in several different ways;
e. Data saving sessions in several locations simultaneously using the "mirroring" functionality.

A basic approach to improving the quality and performance of backup processes is to reduce the amount of data copied and is shown in Table 1.

Thus, the centralized data backup program proposed to be implemented involves the combination of full backup sessions with incremental ones, improving both the execution time of the backup sessions but especially reducing the amount of data to be saved [13]. At the same time, the implementation of the centralized backup system has the advantage that data restoration processes will be simplified and their execution time will be reduced, resulting in an increase in the quality of execution of backup and restoration processes and by reducing the percentage of failed sessions.

Figures 3 and 4 describes the processes related to the data backup system proposed for implementation in an intellectual property organization. Thus, the management of this system has as main objective:

- Determining the data to be saved and especially the frequency of these backup sessions;
- Establishing the data retention period (processes M9, M10 in figure 4);
- Defining backup sessions in accordance with institutional needs;
- Realization of the management of backup media (disk volumes or pools of magnetic tapes LTO6 - processes M2, M3 from figure 3);
- Performing the management of backup sessions performed correctly and those terminated with errors (processes M5, M6, M7, M8 in Figure 4);

- Realization of the supply with storage media (disk volumes or pools of magnetic tapes - processes S1, S2 and S6 from figures 3 and 4);
- Coordination of the support processes necessary for the maintenance and upkeep of the entire centralized backup system (processes S3, S4, S5 and S6 in figures 3 and 4).
- Permanent monitoring of disk volumes and pools of LTO6 magnetic tapes in order to assess the need for supply both in the short term and for a long period of time (more than 6 months), graphically representing the processes S7 and S8 in figure 4.

The catalog database stores information about the following:

a. Backup, restore, object copy, object consolidation, object verification, and support session management sessions. This is a copy of the information sent by the backup session monitoring system (process A13 in Figure 4).
b. Objects to be saved, their versions and copies of the objects. In the case of encrypted object versions, the key and identifiers (KeyID-StoreID) are stored.

Table 1.

Improving the management and performance of backup and restore processes

Backup process	Full Backup	Incremental backup
Resources	It takes longer to complete than incremental backup and requires more media space.	The backup process is performed only for changes made from the previous save session, which requires less time and media space.
Storage media management	If a stand-alone device with a single writing unit is used, a manual backup media change is required if a save session cannot fit on a single media	The system is unlikely to require additional storage media.
Restore	Simplified and fast restoration process.	The restoration process takes longer due to the large number of media required.
Impact on the internal backup database	It takes a lot of space in the database.	It takes less space in the database.

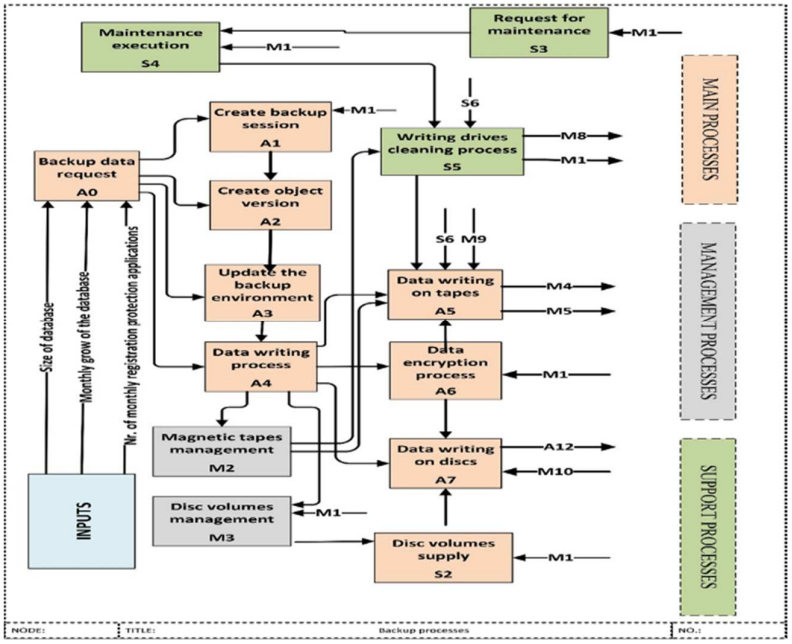

Fig. 3. Management of the backup processes (1st part)

c. Positions of objects copied to media. For each backup object, the central backup system stores information about the backup media and data segments used for backup.

The details part of the binary file catalog (DCBF) part of processes A10, A12 in figure 4 stores information about the following:

a. Names of copied files (file names) along with the name of the client system. The names of the files created between the backup copies are added to the DCBF.
b. Information about backup file versions, their size, modification times, attributes/protection, and positions of backup copies on the backup medium (position on magnetic tapes or disk volumes). Thus, a DC (Detail Catalog) binary file is created for each data protection environment used for backup. When the media is overwritten, the old binary is deleted and a new one is created.

During the data encryption process (process A6 in figure 3) written either on disk volumes (process A7 in figure 3) or on LTO6 magnetic tape support (process A5 in figure 3), all keys created, either manually or automatically, during backup sessions, encrypted files are stored in a keystore file [14]. The keys can also be used for copying objects, checking objects, and during restore sessions. In the case of hardware encryption, they can also be used for object consolidation sessions [15].

In the case of software encryption, the key identifiers (each consisting of a KeyID and a StoreID) are connected to the encrypted object versions. This mapping is stored in the catalog database. Different objects in an environment may have different encryption keys (software). For hardware encryption, key identifiers are mapped to the average ID and these mappings are stored in a catalog file. This file contains the information needed to allow you to export an encrypted medium to a different medium.

When a backup session is started (process A1 in Figure 3), a session record is created in the internal database of the centralized backup system (process A2 in Figure 3).

Also, for each object in the session, a record of the version of the object is created. Both records are stored in this internal database (process A3 in Figure 3) having several attributes. Backup session management updates media files during a backup session (process A5 in Figure 3).

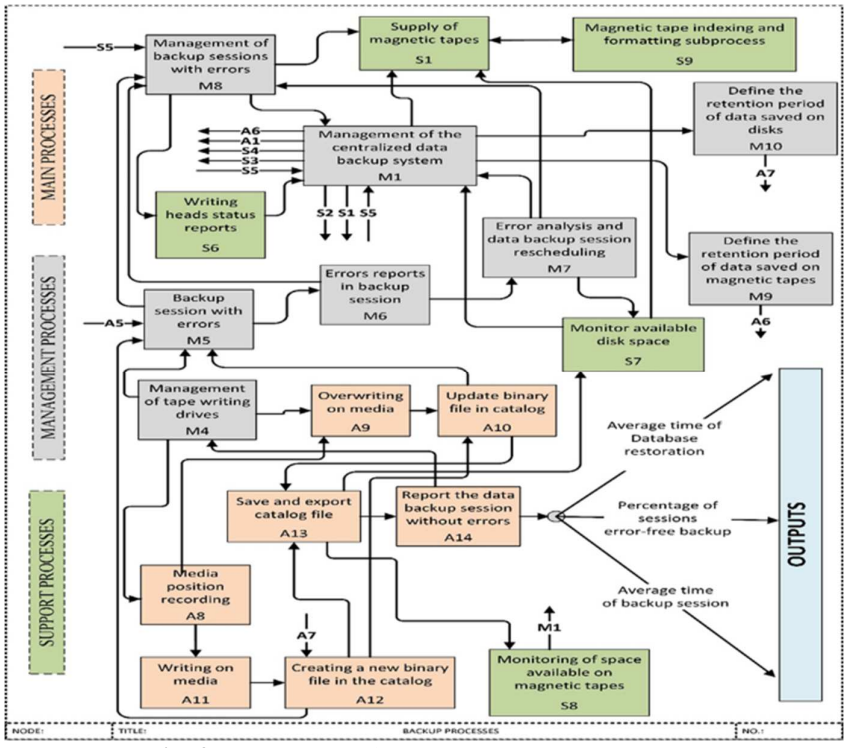

Fig. 4. Management of the backup processes (2nd part)

All media recordings are stored in the MMDB part and are allocated for a backup according to the backup policies established by the management of the centralized backup system (processes M4, M1 in figure 3). When a data segment (and a catalog segment after it) is

written to tape, a record of the media position stored in the internal database for each version of the object that was part of that data segment is recorded. In addition, the catalog is stored in the Detailed Catalog (DC) binary file (process A13 in Figure 4).

The quality of media influences the way media is selected for a backup session, as it affects the ability to write to the environment and read the data contained on it [16].

LTO6 tapes in good conditions are selected for use in centralized backup processes and poor storage media are reported to management as damaged and in need of replacement (M4 process in Figure 3).

During the backup process on LTO6 type magnetic tapes, the management of backup sessions and storage media is performed as follows:

a. magnetic tapes are organized in media groups (pool of tapes). If a magnetic tape in a particular pool of tapes is detected to be of poor quality (caused by the large number of reads/writes made up to that point or is damaged) the centralized backup system will mark the entire pool of tapes as damaged and in need of supply with new bands (processes S1 and S6 in figure 4).
b. in case of errors detected at the level of writing heads on the magnetic tapes, the centralized backup system will report to the management the need to perform their cleaning operations or maintenance works (processes S3, S4, S5 and S6 in figures 3, 4).
c. in case of errors detected at the level of unfinished or interrupted backup sessions due to defects or errors in the system, the affected backup sessions will be marked as affected by errors (processes M5, M6 in figure 4), will be rescheduled by management the centralized backup system also analyzing the causes of their occurrence (process M7 in figure 4).

The reporting mechanisms provide various information on the status of the centralized backup system. The reporting system includes: checking the status of the last save session, copying objects, consolidating or checking backup sessions, reporting systems on the organization's network that are not configured for backup, checking storage consumption, media pools, device status, and more others [17]. In this context, reports and report groups will be configured using any Java-supported web browser. The report groups will allow you to easily manage the reports, schedule the reports in the central report and define the criteria for grouping the reports into report groups, while allowing you to customize the reports by defining the parameters.

3. CONCLUSIONS

In a world in a continuous economic evolution, the volume of data produced increases exponentially, thus raising the problem of digitizing and automatization of processes for saving and securing data from the information systems of organizations. In this context, the constant and rapid increase in the volume of data that needs to be saved and / or recovered is one of the challenges that medium or large level organizations have to meet. Within intellectual property organizations, the monthly increase in the volume of data processed and stored is an important parameter that must be taken into account in the design of new architectures for saving and restoring data in case of need. In this sense, the digitalization of all backup processes but especially of monitoring and reporting are important steps that must be implemented at the organization level. Also, the digitization and automation of these processes will be an important step towards increasing the quality of centralized backup processes on different storage media, but especially to improve information management system, in order to increase the availability, confidentiality, but especially the security of data processed within organizations. Organizations need to increase the performance of existing backup and archiving operations to meet the challenge of unstoppable data growth and limited budgets. Backup and restore operations need to run faster, so more data can be protected every day, and the deduplication technology must be used to aggressively reduce the size of backup sets.

Designing the architecture using backup to discs with deduplication can help improve the quality of the backup processes and simply lead

to replacing old tape-based technologies with new, more expensive but faster volumes disk-based challenges. [18].

Data backup operations on disks with deduplication are not just primary storage or a media server with a deduplication function. Based on the choices of deduplication methods and scaling architecture, a large part of the organization's backup environment can be improved, with consequences such as fast and error-free backups, restricting the time allocated to backup sessions by increasing the transfer speed. and last but not least the decrease in the time required for restoration procedures. All this has a significant impact on increasing the quality, availability, confidentiality and security of the organization's data.

4. REFERENCES

[1].Doina Banciu, Ionut Petre, Mihail Dumitrache, *Electronic system for assessing and analysing digital competences in the context of Knowledge Society,* (2019).

[2].Doina Banciu, Mirelille Rădoi, Stefan Belloiu, *Information Security Awareness in Romanian Public Administration: An Exploratory Case Study,* (2020).

[3].Dedupe, L. F, *The Insider's Guide to Data Deduplication.* CreateSpace Independent Publishing Platform, (2010).

[4].Daehee Kim, Sejun Song, Baek-Young Choi, *Data Deduplication for Data Optimization for Storage and Network Systems.* Springer, (2017).

[5].Nelson, S., *Pro Data Backup and Recovery.* Apress, (2011).

[6].Mihail Țîțu, Constantin Oprean, *Managementul calității în economia și organizația bazate pe cunoștințe,* AGIR, (2008).

[7].Preston, C., *Backup & Recovery Inexpensive Backup Solutions for Open Systems,* O'Reilly, (2007).

[8].Carapola, A, *The Data Center Builder's Bible - Book 1: Defining Your Data Center Requirements: Specifying, Designing, Building and Migrating to New Data Centers,* (2018).

[9].Gaurav Barot, Chintan Mehta, Amij Patel, *Rapid - Hadoop Backup and Recovery solutions,* Knowarth, (2015).

[10].Mohn, C., *Learning Veeam® Backup & Replication for VMware vSphere.* Packt Publishing, (2014).

[11].Bradford, R., *Effective MySQL Backup and Recovery.* ORACLE, (2012).

[12].Robert Freeman, Matthew Hart, *Oracle Database 12c Oracle RMAN Backup and Recovery.* ORACLE, (2016).

[13].Michelle Malcher, Darl Kuhn, *Pro Oracle Database 18c Administration: Manage and Safeguard Your Organization's Data 3rd ed. Edition.* Apress, (2019).

[14].Guise, P. d. *Enterprise Systems Backup and Recovery.* CRC Press, (2017).

[15].Dorian Cougias, E. L. Heiberger, Karsten Koop. *The Backup Book Disaster Recovery from Desktop to Data Center.* Schaser-Vartan Books; 3rd edition, (2003).

[16].Goetz Graefe, Wey Guy, Caetano Sauer. *Instant Recovery with Write-Ahead Logging.* Morgan&Claypool Publishers, (2014).

[17].Chomsiri, T. *Sniffing packets on LAN without ARP spoofing.* ICCIT'08. Third International Conference on Vol. 2. IEEE, (2008).

[18].Glinz, M. *On non-functional requirements.* Requirements Engineering Conference, 2007. RE'07. 15th IEEE International, (2007).

CALITATEA ȘI DIGITALIZAREA MANAGEMENTULUI PROCESULUI DE BACKUP A DATELOR ÎN ORGANIZAȚII CARE AU CA OBIECT DE ACTIVITATE PROTECȚIA PROPRIETĂȚII INTELECTUALE

Rezumat: Lucrarea propusă susține din punct de vedere științific necesitatea implementării tehnologiilor centralizate de stocare a datelor pe discuri și digitizarea proceselor aferente în organizațiile de proprietate intelectuală. Lucrarea abordează problema furnizării de platforme centralizate de stocare a datelor în contextul integrării tuturor sistemelor informaționale fizice sau virtuale ale acestor organizații. Cercetarea se referă la o analiză atât a proceselor de salvare centralizată a acestor date pe discuri, cât și a tehnologiilor de securitate a bazelor de date care pot fi implementate în sistemele informaționale din cadrul acestor organizații. Problematica analizei arhitecturilor și tehnologiilor care pot fi aplicate în domeniul securizării și salvării bazelor de date în sistemele informaționale existente în aceste organizații dar mai ales îmbunătățirea managementului și creșterea calității execuției acestor procese sunt aspecte extrem de relevante analizate și explicate în această cercetare. Cercetarea propusă poate fi considerată un punct de vedere argumentat și documentat într-un domeniu de actualitate.

ȚÎȚU Aurel Mihail, Professor, Lucian Blaga University of Sibiu, mihail.titu@ulbsibiu.ro
MOISESCU Radu, Eng, Ph.D Student, Politehnica University of Bucharest, radu_moisescu@yahoo.com
OPREAN Constantin, Professor, Lucian Blaga University of Sibiu, constantin.oprean@ulbsibiu.ro