# Information Security Issues in a Digital Library Environment: A Literature Review

# Dijital Kütüphane Ortamında Bilgi Güvenliği Sorunları: Literatür  Değerlendirmesi

*Audrey ANDAY*, Enrico FRANCESE*, Hugo C. HUURDEMAN*,*
*Muharrem YILMAZ*, Dydimus ZENGENENE**

## Abstract

*This paper aimed to explore the literature on security issues that digital libraries should consider in managing digital resources. Books on information security and network security were consulted as well as several databases such as ERIC, Ebrary, LISA, Science Direct, EbscoHost, ISI, Google Scholar, ProQuest, Emerald Insight, ACM were searched to understand what particular aspect of information security and privacy in digital libraries exist from 2000 - 2010. Security in digital libraries is an issue of the most important, and should be considered carefully in creating policies and strategic plans of institutions wanting to set up a digital library. This paper focused on the four main streams that concerns security in the digital environment, namely: infrastructure, digital content, users and standards and legal issues. This literature review also built upon previous literature reviews, and is one of the few of its kind in the topic.*

***Keywords:** Information security, Digital libraries, Data protection*

## Öz

*Bu çalışma dijital kütüphanelerin kaynakların yönetiminde göz önünde bulundurması gereken güvenlik sorunlarına ilişkin literatürü ortaya koymayı amaçlamaktadır. Bilgi Güvenliği, Ağ Güvenliği, Kişisel Gizlilik konuları üzerine 2000 - 2010 yıl aralığını kapsayan kitaplar ve makaleler ERIC, Ebrary, LISA, Science Direct, EbscoHost, ISI, Google Akademik, ProQuest, Emerald Insight ve ACM gibi çeşitli veri tabanlarından taranmıştır. İncelenen literatürden elde edilen sonuçlara göre, bilgi güvenliği dijital kütüphaneler için son derece önemli bir konudur ve dijitalleşme sürecinde bir kütüphane, güvenlik politikalarını ve stratejik planlarını dikkatle göz önünde bulundurmalıdır. Bu araştırmada dijital ortamda güvenliği ilgilendiren Altyapı, Dijital İçerik, Kullanıcılar, Standartlar ve Hukuki Konular olmak üzere dört ana madde üzerinde durulmuştur. Bu çalışma ayrıca daha önceki literatür taramalarını da kapsamaktadır.*

***Anahtar sözcükler**: Bilgi güvenliği, Dijital kütüphaneler, Veri koruma*

* Master Students; International Master in Digital Library Learning (aganday@gmail.com), (efrancese@gmail.com), (hugohuurdeman@gmail.com), (yilmaz.muharrem@gmail.com), (dydimus.zengenene@gmail.com)

## Introduction

Society has been increasingly dependent on information technology (IT) for several years now. In this Information Age, millions of users (or participants) access and exchange billions of objects of information content in complex work flow processes (e.g., commerce,  learning, health care). The research community uses computer systems to perform research and to disseminate findings.

Information sharing has been made easier and less expensive by Internet technologies and global networking infrastructures, but availability of such information systems comes at the expenses of higher risks. In the long run, information is not preserved, websites tend to disappear frequently and digital media become obsolete easily and there can be an abuse in the privacy of information. Moreover, the integrity of the systems could be compromised. Access control is often described as rules regulating how participants are allowed to access object and could also be viewed as information flow control because every access results in flow of information between entities (either or both participant and object) (Chen, Choo and Chow, 2006).

The integrity and availability of all these systems have to be protected against a number of threats. Hackers, rival corporations, terrorists and even foreign governments have the motive and capability to carry out sophisticated attacks against computer systems (Patel, Qassim and Wills, 2010). Thus, security mechanisms appropriate for Internet-based, real-world applications should be a prerequisite.

Unless an attack is successful or a system is compromised, security in general, intrusion detection (ID) in particular, is rarely noticed by management. When security fails and the notification is too late, only would managers consider viewing the security issue as visible as their organizational needs. Such crisis would finally open the sense of importance of security in any given system (Goodall, Lutters and Kondoli, 2009).

Dorsish et al. (2004, p.391), mentioned in their paper that "effective security solutions depend […] also on people's ability to understand them and use them as part of their work" .

Moreover, Birnbaum (2004), in his talk shared that in today's information-rich world, digital libraries would play an essential role and will assume central positions of even more significance in pervasive systems. They will not only serve as repositories of knowledge and information, and as the primary mechanism for its retrieval and distribution, but they will also be the focal point for the integration of information and scholarship across all boundaries of application, language, and media. Since they will also inevitably become the target of malicious attack by people seeking unauthorized information, and by terrorists seeking to disrupt the global information infrastructure and the physical infrastructures built upon it, it is both timely and essential to study the cyber security characteristics future digital libraries will have to support.

Furthermore, Tyrväinen in 2005, as cited in Fox and ElSherbiny (2011), considered the security as "an important issue in digital library design. Security weaknesses in digital libraries, coupled with attacks or other types of failures, can lead to confidential information being inappropriately accessed, or loss of integrity of the data stored. These in turn can have a damaging effect on the trust of publishers or other content providers, can cause embarrassment or even economic loss to digital library owners, and can even lead to pain and suffering or other serious problems if urgently needed information is unavailable" (Fox and ElSherbiny, 2011, p.8).

This paper reviews literature about security issues in the digital environment specifically what digital libraries should be aware of in the first place.

## Methodology

The search strategy that was employed for this literature review involved searching printed and online materials. Books on information security and network security were consulted as well as several databases such as ERIC, Ebrary, LISA, Science Direct, EbscoHost, ISI, Google Scholar, ProQuest, Emerald Insight, ACM were searched to understand what particular aspect of information security and privacy in digital libraries exist from 2000 - 2010.

Several keywords used to search catalogues and databases include "*digital libraries AND security", "security in digital libraries", "information security in digital libraries", "threats information security", "wireless security", "database security", "system security ontology library", "security AND libraries", "security in libraries", "privacy in libraries", "information security", "digital content security", "information security AND legal aspects", "information security standards", "information security AND digital library", "data protection law"*

A very broad spectrum of articles that deals with the whole concept of security came out; so we decided to limit the articles to those that pertains to the four main streams that concerns security in the digital environment:

1) Infrastructure - This section focused on the importance of security applied in any system infrastructure that covers securing hardware and software, ensuring network security, and looking into Web vulnerabilities that can distract the smooth flow of communication and transfer of information in a wired or wireless environment.
2) Digital content - This section discussed how important it is to also ensure that digital content are secured in a digital environment and describes some of the steps that can be undertaken in order to recover important data and attain the real purpose of preservation.
3) User information security - This section illustrated some issues pertaining to the terms of security of systems, maintaining the confidentiality of users within a digital library environment i.e. their private information are kept in a trustworthy manner and is not used without their knowledge.

4) Standards and legal issues - This section provided an overview of the development of the different existing standards in ensuring security of any system which can serve as basis for formulation of polices and guide in setting up a system in digital environment.

### *1 - Infrastructure*

According to Lampson (2004), people have been working on computer system security for over thirty years and they have registered notable intellectual success. However, the security risk of millions of deployed computer systems is so high that a determined and competent attacker could destroy most of the information on almost any of these systems or steal it from any system that is connected to a network or even attack millions of systems at once.

Library computers are not safe, they are physically vulnerable to theft, damage and destruction, but, most of all, they are vulnerable to attacks by a host of malware agents which include Trojans, viruses, worms, adware, spyware, pornware, keystroke loggers, password stealers and others (Zimerman, 2009). Hackers, viruses, worms, and trojan horses as external extrusions which libraries should be able to handle (Al-Suqri and Afzal, 2007). Computers are not safe because they have the most popular antivirus software; instead it is more dangerous to believe that one is safe when he/she has antivirus software installed. There are criminals who specialize in targeted attacks, making it more difficult to handle the risk with the traditional antivirus systems (Zimerman, 2009). Given the value of information that they hold, digital libraries have to be worried about this problem.

Danger is a multifaceted threat which faces every computing environment, however there are protection systems that have to be applied but some are too expensive for a library and they only help to minimize but are never perfect (Zimerman, 2009). In a library environment it is even harder since it is difficult to control behavior of many users. Lampson (2004) summarizes it all by his phrase "security is pain," arguing that the threat of IT security does not seem very high until one is attacked, however implementation of security is expensive and takes time from hours of production even if it does not directly contribute to production. The "pain" is even higher in libraries where the output is a service which is usually offered free of charge.

#### *1.1 Securing the Hardware*

Hardware security is the security of such equipment as computers, printers, monitors etc which libraries find indispensible in their day today functions especially in this digital era. There is need to keep such hardware in secure rooms under physical lock and key and an inventory system should be implemented for easy tracking. Control deters theft of property, unauthorized access to servers thereby preventing tampering with server settings, corrupting data, or gaining access to programs and confidential information (National Forum on Education Statistic, 2003). In order to maintain hardware security, it is important to implement strong physical security measures.

*Network security*

In a digital library "...resources are accessed via the Internet and networks are playing a vital role in connecting these information sources" (Singh, 2003). "In the digital age availability of secure, efficient and cost effective networks of access, would be the core competency of the libraries". It would be vital for libraries to secure networks so that the integrity of data can be maintained (Al-Suqri and Afzal, 2007). Network equipment include hubs, routers switches and cabling. For the hardware that supports the network it is necessary to implement security measures that correspond to all other sensitive hardware equipment (National Forum on Education Statistic, 2003). Computer networks now exist as wired and/or wireless networks and security measures in these environments are different. Libraries tend to use wired networks for machines which are fixed in their premises. Wireless networks are used for connecting users who might be having their own mobile gadgets to connect to the network.

To ensure security of physical networks, it is important not to allow users to install unauthorized network equipment, use secure passwords for root access, ensure proper cabling and cable protection (National Forum on Education Statistic, 2003).

A wireless network is "…a network that uses high-frequency radio waves rather than wire to communicate between two nodes." The wireless network infrastructure has brought about better flexibilities in terms of geographical limitation as well as hardware and software accommodated. Mobile phones and other gadgets other than personal computers are joining the network realms which were previously the domain of personal computers connected through wires (Khalil, 2004).

Wireless networks will be the standard mode for information access for both on-campus and classrooms connectivity. This technology is already helping students to interact with digital library systems on the net (Khalil, 2004). Wireless networks have however also brought with them a great degree of risks as far as network security is concerned. Unlike in the wired network, security in a wireless network is more of concern because network transmissions are available to anyone within the transmitter with the appropriate antenna, physical access controls like doors and locks do not help. Sniffing (intercepting) is much easier because the radio transmissions are designed to be processed by any receiver within the range and also that they have funny boundaries beyond the intended one (Gast, 2002). For that reason, the wireless networks is a double edged sword which possesses both high potential and high risks (Porter, 2002, p.16). Wired networks are also insecure since it is possible for an attacker to tap electromagnetic energy that is radiated by wired networks; however this is by use of sophisticated equipment and involves relative proximity to the cables unlike the wireless signal which can easily overlap across the intended boundaries (Porter, 2002, p.21). Due to the increase on the use of mobile gadgets, digital libraries are increasingly being accessed via wireless networks. That implies the need to consider investment in wireless network security if the integrity of information resources is to be maintained.

### *1.2 Operating system security*

The operating system is the underlying system on which application programmes run. Therefore, the choice of an operating system plays a critical role in ensuring system security. Operating systems ensure access to centralized resources including applications, access privileges can be granted or restricted thereby regulating the use of network resources. Some operating systems are easier to run yet they are less secure than those that might be difficult to run. In any case the system must be hardened or secured by removing unnecessary functions, restricting access and tracking changes and processes. There are several free open source operating systems available for free and proprietary operating systems for which libraries have to pay; however the cost of purchasing a system is not a guarantee for security. It is however possible to run a mixed computing environment where systems run on different operating systems but there is need for experience and high degree of expertise in administering such environments even though they guarantee better security (National Forum on Education Statistic, 2003). Libraries are therefore advised to consider establishing mixed computing environments even if the costs of maintenance are high.

### *Database security*

Databases are very critical parts of the library information system as the key hosts of metadata, and other administrative information. Databases employ security systems as those of operating systems but users are assigned certain types of groups called roles. For example the head librarian and the library clerk have different roles in the system and that controls what each user can view or edit in the database. Database security can be maintained discretely or can be integrated with operating systems. That implies that users will require only one logon into the system. Database security mechanisms are effective if they are used in conjunction with proper security mechanisms implemented at the front end application like dynamic web pages (National Forum on Education Statistic, 2003). Databases have the capability to offer access to resources as defined by roles and profiles and should be based on the respective functions. A database should also have tracking features that can track when the database was accessed by whom and what changes took place. For instance; it must be possible to trace who added an article to the collection and when. Data transmission should be secured using protocols such as Secure Socket Later (SSL) or Secure Shell (SSH).

SSL is "a public key cryptography based confidentiality mechanism…" which is historically associated with web pages accessed via the secure hypertext transfer protocol (https) even though it can be used to encapsulate any protocol. Porter (2002) judges that SSL is best for protecting transaction based protocols such as web traffic and mail transactions. SSH is a secure replacement for commands such as rlogin, rcmd, and rshel. SSH also uses public key cryptography like SSL but does not rely on trusted authority to issue the public/private key pairs (Porter, 2002).

### *1.3 Web application level vulnerabilities*

Despite the laws in European countries that mandate secure sites, many library websites have serious security flaws which render then vulnerable to attacks (Kuzma, 2010). From a research conducted in European countries, almost 80 percent of web related flaws were caused by web application vulnerabilities with the three main common types being: Cross scripting, Denial of Service and SQL injection. Major causes for these problems are pointed to be, lack of updating software versions, developers install the default software and forget the need to update, lack of consideration of security flaws, lack of upgrading software correctly and lack of effecting coding practice during designing and development (Kuzma, 2010).

#### *Cross-site Scripting*

Cross-site scripting is a security vulnerability that allows the injection of programming code by malicious third parties into web pages hosted on a server. This allows risks by allowing fishers or fraudsters to launch an attack without directly targeting or gaining access to a legitimate website. This allows unknowing and unsuspecting web visitors to see forms input and send data or to be exposed to malicious downloads on other content while viewing your website (Cyveillance, 2008).

#### *Denial of Service*

Denial of Service (DoS) is a type of attack that prevents access to network resources and this can be devastating and difficult to protect against and DoS "…involves flooding the network with traffic choking the transmission lines and preventing other legitimate users from accessing services on the network". Denial of access can come as various types of attacks at different layers of the OSI model but all leading to network flooding (Porter, 2002).

#### *SQL Injection*

"SQL injection attacks are a type of injection attack, in which SQL commands are injected into data-plane input, in order to affect the execution of predefined SQL statements. It is a common threat in web applications that lack proper sanitization on user-supplied input used in SQL queries" (Guimarães, 2009).

Due to the increased need to offer computer aided web-based services, libraries must be aware of all these possible infrastructural threats and protect their data and the entire system.

## ***2 - Data Security***

The core of any information system is the data contained in it: Libraries make no exception. With "data" we mean both the documents and the meta-information applied to them: OCLC (2006) reminds that "For disaster prevention and recovery, all data (content and metadata) is considered of equal value". A secured system with corrupted data is useless; in the same way, the data storage within a frail infrastructure is weak and exposed to danger.

System security and data protection go hand in hand, part of the same side in what Fox (2006) calls the "two-front war". From one side we must protect our patrons, their privacy and confidentiality as well as their electronic devices. On the other hand there is the need to protect the digital content itself and the electronic infrastructure from abuse.

### *2.1 Background Fears*

The risks of digital preservation which libraries and archival institutions have to consider were foreseen since the end of the 1990s. Authors called for attention to these issues assuming dramatic tones, speaking of "digital dark ages" (Kuny, 1997) and "a time-bomb for digital libraries" (Hedstrom, 1998).

The first author to claim risks for the digital preservation was Rothenberg (1995, p.2), who indicated the risk of format obsolescence as the main threat to the digital cultural heritage: "although its reproducibility make digital information theoretically invulnerable to the ravages of time, the physical media on which it is stored are far from eternal. [...] The contents of most digital media [...] become unusably obsolete much sooner, as they are superseded by new media or incompatible formats" .

Moving his steps from this article, Kuny raised several points which in part repeat the concerns pointed out by Rothenberg (1995):

◊ Enormous amounts of digital information are already lost forever.

◊ Information technologies become obsolete very quickly.

◊ Document and media formats continue to proliferate.

◊ Technology standards will not solve fundamental issues in the preservation of digital information.

◊ Libraries will shortly see a demographic bulge of electronic material as the baby boom generation of authors and academics contribute material gathered during their careers.

◊ Much material will never make it into library collections for preservation because of increasingly restrictive intellectual property and licensing regimes.

◊ Archiving and preservation functions in a digital environment will increasingly become privatized as information continues to be commodified.

Kuny shared two main concerns with Rothenberg (1995): The actual risk of obsolescence and the distrust for standards. The lack of faith in standards is due to the fact that commercial software vendors are not willing to play on. "The challenge in preserving electronic information is not primarily a technological one, it is a sociological one" (Kuny, 1997, p.4).

Hedstrom (1998) looked at the standard issue with a different perspective. For her, the problem is that "Digital preservation is constrained by the absence of established standards, protocols". In 1998 she just saw the situation as not mature enough.

We will see in section 4 of this paper that since the time of Hedstrom and Kuny were writing, a lot of new developments happened in the field of standards. In this section we focus on the problems of data safety and preservation.

### *2.2 Obsolescence*

A definition of obsolescence is given by Pearson (2008). Reminding that "A file format is a particular way to encode information for storage and use", he defines "obsolescence" as: "the development of new format encodings that take the place of already existing formats in the marketplace of use; and the changes in the availability of presentation tools, generally (although not exclusively) in the direction of decreasing availability, for any particular file format" (Pearson, 2008, p.91).

According to Rosenthal (2010a), obsolescence has proved to be a minor risk: "format obsolescence is a rare problem that happens infrequently to a minority of unpopular formats". Nevertheless, he proposes two solutions: a standard solution and an alternative one.

The standard solution consists of migration: it is based upon public registries of format specifications and the creation of software which converts files in obsolete formats to usable files. A format registry is "a repository for format representation information or, in other words, descriptive, administrative, and technical metadata about digital formats, including the definition of the syntactic and semantic characteristics of the registered formats. This metadata defines the significant properties of digital formats with regard to the long-term preservation of digital objects" (Abrams, 2005, p.131).

The alternate model is based upon emulation: the obsolete file is rendered in a replica of his original environment. Open-source technology is very important in order to create working emulators (Rosenthal, 2010b). Abrams (2005, p.129) notes that emulation differs from migration in the fact that the file is not manipulated but its integrity is kept at its original conditions.

### *2.3 Data security backup*

The main safety measure for the integrity of data is the backup (Whitman, 2003). Hadow (2009) clearly indicates backup as the main way to protect content ("The most reliable backups store the copied data off the premises, preserving it from physical damage.").

The OCLC Digital Archive Preservation Policy (OCLC, 2006, p.10) details a state-of-the-art backup strategy. The main points can be summarized as: backups are made on tape support; data and metadata are treated together; operations are handled by specialized dedicated staff ("OCLC maintains staff solely dedicated to network and system security, including at least one Certified Information Systems Security

Professional.”); backups are kept in secure off-site storage facilities (“All computer rooms are protected from fire by a halon gas fire suppression system. All computer rooms are climate-controlled with raised-floor environments”) whose access is strictly regulated (“Access privileges to the computer room are limited and are reviewed every three months. Each access is logged, recording information such as the staff person entering, the door entered, and the time”). The off-site facilities must “meet the highest industry standards for safety and security”.

Rosenthal (2010b) shows how data storage has become easier and easier in the last years thanks to the development of technology and the lowering of the related costs: “Storage is cheap, so if there is a chance the data could possibly be useful, we keep it. We know that storage isn’t completely reliable, so we keep backup copies as well.” Despite this, backup is not 100% reliable and easy to achieve. In the same article Rosenthal shows the difficulties related to the planning and cost of the backup systems: “Our inability to compute how many backup copies we need to achieve a reliability target is something we are just going to have to live with.” He also reminds that “in the real world failures are inevitable”, especially in the large-scale digital preservation projects required by today’s institutions.

In a different article of the same year Rosenthal (2010a) returns to the false claim that storage is free or low cost: “again, at the scale of real digital preservation and with an appropriate number of copies this is certainly not true”. Then “it is often said that “bit preservation is a solved problem”, but at the scales and for the durations needed in digital preservation this is unfortunately not the case”.

In the backup era, this solution has a drawback: “Ironically that ability to mirror and duplicate digital objects also becomes a liability when data is stolen that was not intended for public consumption” (Fox, 2006). This aspect of digital information is seen as a potential flaw even by Kuny (1997), who notices how “Digital collections facilitate access, but do not facilitate preservation”, and by Hedstrom (1998) who claims that “The two terms “mass storage” and “long-term preservation” embody a contradiction in the current state of affairs of digital library development, representing a time bomb that threatens the long-term viability”.

Maniatis et al. (2005) also point out the peculiarity of the backup strategies required by digital preservation projects. They make three starting points: “Digital preservation systems have some unusual features. First, such systems must be very cheap to build and maintain, which precludes high-performance hardware such as RAID (Patterson et al. 1988 as cited by Maniatis, 2005) or complicated administration. Second, they need not operate quickly. Their purpose is to prevent rather than expedite change to data. Third, without central control and in the face of possible interference from attackers or catastrophic failures of storage media such as fire or theft, must function properly for decades.”

### *2.4 Policies*

Data backups, just like all the security measures discussed in section 1, must be part of what we call "disaster recovery plan". Fox puts it very plain: "Having a disaster recovery plan is very important" (2006, p.255). All the literature agreed that backup practice is nothing if not supported by a clear preservation policy, which involves also security and information literacy and staff training and education (Kouzma, 2010, and Balas, 2005). Whitman (2003) and Parkin (2009) also stress on the importance of policies for an efficient digital preservation plan.

We examined two policies: "OCLC Digital Archive Preservation Policy and Supporting Documentation" (OCLC, 2006) and the report "Digital Preservation Policies", prepared for JISC in 2008 (Beagrie, Semple, Williams and Wright, 2008).

OCLC openly confirms the claims reported by Kouzma (2010, p.5): "A preservation strategy must include more than just what can be achieved by good system back-up procedures. A strategy is needed also to ensure the long-term accessibility of digital content objects deemed to have enduring value" .

The JISC report is interesting also because it aims to be a model for further preservation projects. "Our objective therefore has been to produce a practical guide for developing an institutional digital preservation policy" (Beagrie et.al, 2008).

The policies addressed the preservation problem in an organic way, embracing all the aspects of the institution: from the definition of the Principle Statement which guide the policy itself, to the connection with all other practices engaged by the institute, to the definition of the content and the practices to implement.

## ***3 - User Information Security***

Computer systems have become an essential element of libraries. As patrons are using library systems, a large amount of transaction data about users is being recorded, and often stored in the systems. This development has severe implications for the security of user data. The already mentioned "two-front war" that is being fought by libraries results in the need to protect library systems against various types of abuse (see section 1), and the need to guard the confidentiality of their users (Fox, 2006, p.250). Not only hackers and criminals can try to gather confidential data, but also government agencies can do inquiries about library users (Bowers, 2006).

The previous section covered the safety of data with respect to accidents and obsolescence, and data preservation policies. Another important issue in libraries is the security of user information, which will be discussed in this section. We will firstly define privacy and confidentiality. Subsequently, we discuss types of privacy issues in the library context, as gathered from the literature. We will also discuss threats to security of user information, trust issues and finally look at security principles regarding user information used by libraries.

### *3.1 Definition of privacy and confidentiality*

The term ‘privacy’ is closely related to the security of user information. Bowers has defined privacy as “information about an individual that is unavailable to others” (Bowers, 2006, p.377). According to Gorman, privacy is one of the “core values” of a library, so it is of key importance for libraries to assure the privacy of their patrons. Gorman also discusses the notions of “spatial privacy” and “information privacy”. Spatial privacy is the right to be where you want to be (and with whom), while information privacy is the right to control your personal information, without being monitored by external entities (Gorman, 2001, p.2).

Closely related to privacy is confidentiality, which is the right of library users to keep the nature of their research and their library transactions private. In the delivery of library services, the right of patrons to have the nature of their research and library transactions remain private. Under the guidance of the *ALA Code of Ethics*, librarians and library staff members are encouraged to (American Library Association, 2008) “protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted.”

### *3.2 Privacy issues*

Privacy and confidentiality issues in libraries are connected to different aspects of library use. For example, to perform searches in the database, patrons are using library systems. So there are ‘transactions’ between libraries and library users (Gorman, 2001). These transactions could be monitored and stored by libraries. As Newby (2002, p.3) indicates: “by necessity, these are linked to identifying information for individual patrons who borrow books or other materials”.

Some of this user-related information is vital for running the library (for example, which books are lent by a certain patron at this moment), while other information is not necessarily needed for normal operations (for example, the complete lending history of library users). Gorman (2001, p.5), describes this as “accumulation of personal data for legal purposes”, which can have “abuse potential”. Also, in the age of computerization, more confidential information could be registered. “Users’ names, addresses, website logs, e-mail logs are archived by the computerised management system of the libraries automatically” (Al-Suqri and Afzal, 2007, p.45). Neuhaus (2003) also mentions that chat and e-mail-based chat reference services can be captured on the library’s or third-party servers, a “digital trail” that could lead to potential privacy issues (p.10). All in all, the more information is stored, the greater the potential problem becomes if a library system’s security is compromised.

The ethical dimension of privacy is described by Peters (1999, as cited in Neuhaus, 2003, p.5-6), who mentions three broad ways of unethical ways of data collection. First of all, the data itself might be private or confidential, “so the mere collection of the data represents an invasion of privacy”. Also, “the methods to collect the data might be unethical” and thirdly, “the use of data may be unethical”.

### *3.3 Threats to confidentiality of user information*

The dangers that libraries face with regards to the confidentiality of user information can have different natures. The main external threat categories, as gathered from the literature, are computer systems security breaches, social engineering attacks and government agencies' enquiries.

#### *Computer systems security breaches*

As described in section 1, computer systems of libraries can be vulnerable to security problems of different nature. This means that the user's privacy is at risk, because others can gain access to his or her personal data and behavior statistics (Whitman, 2003). Despite these potential problems, privacy related features "are seldom given more than routine attention" when libraries select systems from suppliers of software systems (Sturges et al., 2003, p.6).

#### *Social engineering*

Another way to get unauthorized access to user data is by means of 'social engineering', which is the use of non-technical means to gain unauthorized access to information or computer systems (a "non-technical hack"). Social engineering works mainly because most people work under the assumption that others are essentially honest. Libraries are vulnerable to social engineering, because of "ignorance" (lack of concern) and "institutional psychology" (librarians are expected to fulfill patrons' requests without questions or bias) (Thompson, 2006, p.3).

Sturges et al. (2003, p.6) also indicate that "many librarians are unaware of the procedures that they should follow when they receive an enquiry concerning personal data". This indicates that libraries are potentially susceptible to the threats of social engineering, as also put forward by Anderson (2007, as cited in Dlaminia, Eloffa and Eloffb (2009, p.9): "most of today's security challenges are to a greater extent related to the human and organisational aspects of security".

#### *Government agencies' enquiries*

In some countries, government organizations have legal rights to access user information from libraries. In the United States, the "Patriot Act", issued after the terrorist attacks on September 11th 2001, allows government bodies to access user information that is gathered by libraries more easily (Bowers, 2006). Even the number of government requests for information by government agencies in the USA is highly confidential (Sutlieff and Chelin, 2010). This is a major issue, as Fox also indicates: "we now have an unprecedented situation in which librarians are pitted against government agencies in a battle over patron privacy and national security (Fox, 2006, p.250). However, in Neuhaus' (2003, p.9) article, this risk is downplayed, as he indicated that in the aftermath of the terrorists attacks on September 11, 2001, "limited information indicated that the FBI primarily sought computer and e-mail records of library users".

### *3.4 Trust issues*

The threats to user information security discussed above could lead to trust issues. In general, library users seem to trust their libraries. Sturges et al. (2006) conducted a survey among library users, that showed that 89% of all users expressed no concern about their privacy when using library systems. The question is whether this trust is deserved or not. Neuhaus (2003, p.26) argues that "the library community possesses a decade-long tradition of trying to protect the privacy and confidentiality of library users". However, a survey by Sturges et al. (2006) among librarians contradicts this: only 14% of all surveyed institutions have a privacy policy, and 64% of libraries have a data protection policy. As Sturges et al. (2006, p.6) put it: "there exists a genuine gap between the privacy protection that the users believe they can expect from the library, and the preparedness of librarians for providing this protection.".

This neglect of data protection could have severe implications if the privacy of users is not safeguarded, as summed up by Bowers (2006, p.377):

"If a person does not have an expectation that their library records will be kept confidential, they may be unwilling to ask questions, perform a search, read a book on the premises, or check out a book on a controversial subject for fear of judgment by the community they live in or society at large, or for fear of retribution by the government".

### *3.5 Library privacy plans and policies*

In order to maintain user trust, libraries should take measures to protect their patron's confidential data. Gorman (2001, p.10) indicates that this can be done in the form of a privacy plan, which should be built on a combination of 'principle' (the natural law right to privacy) and 'experience' (the illumination and exemplification of a principle).

Different authors offer recommendations for libraries regarding privacy plans. Newby proposes a set of recommendations: (Newby, 2002, p.3)

1. Maintain a comprehensive list of data that may be collected and the circumstances
2. For each type of data, what risks of misuse exist?
3. Specify a policy for the collection of data and possible misuses.
4. Identify personnel responsible for ensuring the policies are followed, and for remediation as needed.

So these recommendations consist of listing sensitive data and the risks, devising a data collection policy and the training of personnel.

Neuhaus (2003) discusses a number of policies to limit privacy risks, that are related to Newby's third recommendation above. First of all, libraries should use anonymity in reference software as much as possible. A second step is the limitation of collection of personal information to what is needed for a specific transaction. A third step mentioned by Neuhaus is the restriction of the number of persons who have access to certain records.

The fourth recommendation point of Newby handles about the personnel of a library. By identifying responsible personnel, you can find out whom to train and educate, though Newby does not mention how to do this. Whitman (2003, p.3) points out that Security Education, Training and Awareness (SETA) programs are a possible way to train and educate library personnel. Using these training programs, they can acquire the necessary skills and knowledge to effectively deal with confidential information.

Also, it is important to communicate the privacy plans of libraries to the outside world, in the form of privacy policies. This way, users can find out what information libraries collect, who has access to the information, whether the library uses encryption, etc. (Neuhaus, 2003, p.10). Privacy policies are also influenced by national regulations. It is possible to base these principles on general principles, for example the International Safe Harbor privacy principles (Gorman, 2001).

It might be valuable for libraries to consider using privacy principles in their organization, as Sturges' concludes: "The knowledge, plans and procedures to deliver data protection and confidentiality in practice are not present in libraries generally" (Sturges et al., 2003, p.6). This problem can lead to legal issues, so adjusting security practices to national and international standards might be a good idea for libraries. This is discussed in the next section.

### *4 - Standards and Legal Issues*

From the discussion above, it emerges that, organizations including libraries need to be fully aware of the need to devote more resources to the protection of information assets, and that information security must become a top concern in management (The Government of the Hong Kong Special Administrative Region, 2008). Information, in its various forms, is arguably the most important asset" (Gerber and Solms, 2008, p.124). It is important in a sense that it helps people in their quest for further information and higher knowledge which when processed accordingly can lead to sound decision making and proper management of tasks and challenges. This section of the paper includes, in brief, information about usually adopted standards and regulations for information security.

The corpus of laws related to data security is complex: "Data security management is just as much about employment law, property law and contract law as it is about data protection law" (Turle, 2009, p.54). Smedinghoff (2004) goes further in illustrating this concept: "Some laws seek to protect the company and its shareholders, investors, and business partners. Others focus on the interests of individual employees, customers, and prospects. And in other cases, governmental regulatory interests, or evidentiary requirements are at stake. Many of the requirements are industry-specific (e.g., focused on the financial industry or the health care industry) or data-specific (e.g., focused on personal information or financial data). Others focus only on public companies" (Smedinghoff, 2004, p.3). Multiple sides of the rights sphere are intertwined in participating in the data security matter.

Information security plays an important role in protecting the assets of an organisation. As no single formula can guarantee 100% security, there is a need for a set of benchmarks or standards to help ensure that an adequate level of security is attained, resources are used efficiently, and the best security practices are adopted (HKSAR, 2008, p.3). There are various standards and regulations that are available for information security. Mainly ISO standards dominated the in the field. Besides ISO standards, some non-ISO standards are also in use.

*4.1 ISO Standards*

The ISO/IEC 27000-series numbering "ISO27k" has been reserved for a family of information security management standards derived from British Standard BS 7799 (ISO Standards, 2007). In this section, two of the main standards for information security are discussed: ISO/IEC 27002:2005 and ISO/IEC 27001:200.

*ISO/IEC 27002:2005 (Code of Practice for Information Security Management)*

ISO/IEC 27002:2005 has developed from BS7799, published in the mid-1990's. The British Standard was adopted by ISO/IEC as ISO/IEC 17799:2000, revised in 2005, and renumbered (but otherwise unchanged) in 2007 to align with the other ISO/IEC 27000-series standards.

This standard contains guidelines and best practices recommendations for these 10 security domains: security policy; organisation of information security; asset management; human resources security; physical and environmental security; communications and operations management; access control; information systems acquisition, development and maintenance; information security incident management; business continuity management; and compliance (HKSAR, 2008, p.5). This model for determining legal requirements will form part of a superior model that will assist in determining all information security requirements.

*ISO/IEC 27001:2005 (Information Security Management System - Requirements)*

The origin of the standard comes from BS 7799 as well. The first part, containing the best practices for Information Security Management, was revised in 1998; after a lengthy discussion in the worldwide standards bodies, it was eventually adopted by ISO as ISO/IEC 17799, "Information Technology - Code of practice for information security management" in 2000. ISO/IEC 17799 was then revised in June 2005 and finally incorporated in the ISO 27000 series of standards as ISO/IEC 27002 in July 2007 (HKSAR, 2008, p.5). It specifies the requirements for establishing, implementing, operating, monitoring, reviewing, maintaining and improving a documented Information Security Management System (ISMS) within an organisation (Ozkan and Karabacak, 2010, p.567).

BS 7799 was a standard originally published by BSI Group in 1995. It was written by the United Kingdom Government's Department of Trade and Industry (DTI), and consisted of several parts. This standard offers the PDCA model:

a) “Plan” phase - establishing the ISMS
b) “Do” phase - implementing and operating the ISMS
c) “Check” phase - monitoring and reviewing the ISMS
d) “Act” phase - maintaining and improving the ISMS

### *4.2 Non-ISO Standards*

There are some other commonly used information security standards, which are not under the ISO body of standards. The OECD (Organization for Economic Cooperation and Development) Guidelines for the Security of Information Systems and Networks, GAISP Generally Accepted Information Security Practices, Payment Card Industry Data Security Standard and COBIT (The Control Objectives for Information and related Technology) are some of the common information related standards as non-ISO Standards. These standards are mainly comprised of security issues on the national as well as international level. Many ISO standards actually originated from these standards. For example BS 7799 from the British Standards Institute (BSI) served as a basis for ISO/IEC 27001 and ISO/IEC 27002, as mentioned in the previous section. In addition to the BSI, there are some other organizations which create and publish information-related standards besides ISO, like the Bundesamt fur Sicherheit in der Informationstechnik (Germany) and the Information Security Forum (ISF).

The British Standards Institute (United Kingdom) publishes standards related to information security. “Information technology security techniques” is the most known one and contains many sub standards under the title (BSI, 2010).

The “Bundesamt für Sicherheit in der Informationstechnik” (Germany), in English “Federal Office for Information Security” has established standards for the development of IT security concepts. It is the German government agency in charge of managing computer and communication security for the German government.

ISF Information Security Forum (United Kingdom) is an international, independent, non-profit organization dedicated to benchmarking and identifying good practice in information security. For almost 300 major organizations, including half of the Global 500 and many of the world’s largest corporations and public sector organizations, the Information Security Forum (ISF) plays a significant role in the fight against growing threats to their information (Jones, 2006).

The OECD Working Party on Information Security and Privacy (WPISP) promotes an internationally coordinated approach to policymaking in security and protection of privacy and personal data in order to help build trust in the global information society and facilitate electronic commerce. The OECD Guidelines for the Security of Information Systems and Networks (WPISP), the nine principles set out in the Guidelines reflect the agreed position of OECD countries and have the support of the international business community and consumer protection organizations (Information Security and Privacy, 2010).

The Information Systems Security Association (ISSA) is a not-for-profit, international organization of information security professionals and practitioners. ISSA released GAISP (Generally Accepted Information Security Practices) which has been formed to address the fact that the practice of Information Security (IS) is now at a crossroads. While IS professionals have made strides towards staying a step ahead of rapidly evolving threats, in moving from reactive to proactive practices the professional community has grown in a fragmented manner with little cohesive organization (ISSA, 2004).

## Conclusion

This paper has reviewed the literature on information security which librarians have to consider in setting up and managing digital libraries.

The literature revealed that maintaining a secure infrastructure is necessary, however it implies costs. The need for libraries to have network security has been stressed as a necessity in digital libraries. Security risks, especially web application vulnerabilities, are not taken seriously in libraries, when most of the causes of such vulnerabilities are negligent human errors. Generally speaking, an investment in security literacy is recommended for libraries.

The rising amount of digital information, in combination with rapid developments in technologies and formats, makes it very hard to preserve the digital data. Solutions for obsolescence are migration and emulation. The literature indicates that backups are the most common way to prevent data loss. However, only backup is not enough, but it is also important for libraries to have clear preservation policies as well.

Security issues do not concern only the technological infrastructure or the data, but also affect the users' trust and behavior. The increasing use of computer technology has resulted in an increase in collection of personal data, which could lead to potential privacy and confidentiality issues. Different threats to user information exist, as the security of computers system can be breached in different ways, and also government agencies can do inquiries about library users. As the literature illustrated, libraries are not always paying much attention to the protection of user data. Therefore it is important for libraries to invest in security literacy. They should also devise privacy plans, to ensure that library patrons can use their systems and services without fear.

Finally, given the proliferation of digital libraries and the influence that they are increasingly having in research and learning, it is imperative that libraries consider taking security issues seriously in order to ensure that their resources and user privacy are secured. To ensure that there is an adequate level of information security, different standards and benchmarks must be used. Our literature review has discussed various ISO and non-ISO standards, which could be used by libraries to review their security policies.

## References

Abrams, S.L. (2005). Establishing a global digital format registry. *Library Trends*, *54*(1), 125-143. doi: 10.1353/lib.2006.0001.

Al-Suqri M. and  Afzal W. (2007). Digital age: Challenges for libraries. *Information, Society and Justice*. *1*(1), 43-48. doi: 10.3734/isj.2007.1105.

American Library Association (2008). *Code of ethics of the American Library Association.* Retrieved on 14th April 2011 from http://www.ala.org/advocacy/proethics/codeofethics/codeethics

Balas, J. (2005). Close the gate, lock the windows, bolt the doors: Securing library computers. *Computers in Libraries*, *(March)*, 28-31.

Beagrie, N., Semple, N., Williams, P. and Wright, R. (2008). *Digital preservation policies part 1: Final report October 2008*. *Strategies*. pp.1-60. Retrieved on 15th April 2011 from http://www.jisc.hosting.eduserv.org.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf.

Birnbaum, J.S. (2004). Cybersecurity considerations for digital libraries in an era of pervasive computing.  In *Proceedings ACM/IEEE Conference on Digital Libraries (JCDL'04)* (pp.169-169) New York: ACM.

Bowers, S. (2006). Privacy and library records. *The Journal of Academic Librarianship, 32*(4), 377-383.

BSI - The British Standards Institute and British Standards Publications (2010). *The British Standards Institute and British Standards Publications*. Retrieved on 5th May 2011 from http://www.standardsuk.com/bsi/.

Chen, S.S., Choo, C.Y. and Chow, R.Y. (2006). Internet security: A novel role/object-based access control for digital libraries. *Journal of Organizational Computing and Electronic Commerce. 16*(2), 87-103.

Cyveillance (2008). *Phishing using cross-site scripting: Definition, illustration and prevention.* Syveillance, USA.

Dlaminia M., Eloffa J. and Eloffb, M. (2009). Information security: The moving target. *Computers & Security, 28*(3-4),189-199.

Dourish, P. et.al. (2004). Security in the wild: User strategies for managing security as an everyday practical problem. *Pers Ubiquit Comput, 8,* 391-401.

Fox, E. and ElSherbiny, N. (2011). *Security and digital libraries, digital libraries - methods and applications*, Kuo Hung Huang (Ed.), InTech,  Retrieved on April 2011 from http://www.intechopen.com/articles/show/title/security-and-digital-libraries

Fox, R. (2006). Vandals at the gates. *OCLC Systems & Services*, *22*(4), 249-255. doi: 10.1108/10650750610706961.

Gast, M. (2002). 802.11 Wireless networks: The definitive guide. O'Reilly:  North Sebastopol, USA

Gerber, M. and  Solms, R. (2008). Information security requirements: Interpreting the legal aspects. *Computers & Security, 27*, 124-135.

Goodall, J.R., Lutters, W. G., and Komlodi, A. (2009). Developing expertise for network intrusion detection. *Information Technology & People*, *22*(2), 92-108.

Gorman, M. (2001). Privacy in the digital environment-issues for libraries. In Libraries and Librarians: Making a Difference in the Knowledge Age. IFLA Council and General Conference. (pp.57-65). Boston, Massachusetts: IFLA

The Government of the Hong Kong Special Administrative Region (2008). *An Overview of Information Security Standards*. Retrieved on 25th March 2011 from www.infosec.gov.hk/english/technical/files/overview.pdf.

Guimarães, Bernardo Damele Assumpção (2009). *Advanced SQL injection to operating system full control. Black Hat Briefings Europe, Amsterdam.* Retrieved on 17th April 2011 from http://www.blackhat.com/presentations/bh-europe-09/Guimaraes/Blackhat-europe-09-Damele-SQLInjection-slides.pdf.

Hadow, K. (2009). Data security for libraries: Prevent problems, don't detect them. *Feliciter*, *55*(2).

Hedstrom, M. (1998). Digital preservation: A time bomb for digital libraries. *Computers and Humanities*, *31*, 189-202.

Information Security and Privacy (2010). *Information security and privacy*. Retrieved on 25th April 2011 from http://www.oecd.org/department/0,3355,en_2649_34255_1_1_1_1_1,00.html.

ISSA (2004). *Generally Accepted Information Security Principles* (GAISP), (3rd ed). Information System Security Association.

ISO Standards (2007). ISO/IEC 27001:2005 (Information Security Management System - Requirements). Retrieved on 2nd April 2011 from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=42103.

Jones, A. (2006). The Information Security Forum. *Infosecurity Today*, *3*(6), 38- 40.

Kuny, T. (1997). *A digital dark ages?* Proceedings of the 63rd IFLA General Conference. International Federation of Library Associations and Institutions. Retrieved on 14th March 2011 from http://www.ifla.org/IV/ifla63/63kuny1.pdf.

Kuzma, J. (2010). European digital libraries: Web security vulnerabilities. *Library Hi Tech*, *28*(3), 402-413. doi: 10.1108/07378831011076657.

Khalil, M.A. (2004). Vision to reality: Applications of wireless laptops in accessing information from digital libraries: End user's view points. *Library Hi-Tech News, 21*(7), 25-29.

Lampson, B. (2004). Computers security in the real world. *Computer, 37*(6), 37-46.

Maniatis, P., Roussopoulos, M., Giuli, T., Rosenthal, D.S.H. and Baker, M. (2005). The LOCKSS Peer-to-peer digital preservation system. *ACM Transactions on Computer Systems, 23*(1).

National Forum on Education Statistics (2003). *Weaving a secure web around education: A guide to technology standards and security*. National Forum on Education Statistics (ED/OERI):Washington DC.

Neuhaus, P. (2003). Privacy and confidentiality in digital reference. *Reference & User Services Quarterly*. *32*(1).

Newby, G. (2000). *Information security in libraries*. Proceedings of the Information Resources Management Association (IRMA) Annual Conference. Hershey, Pennsylvania: IRMA. Anchorage, Alaska. May 20-24, 2000. 6p. Republished in Kisielnicki, Jerzy (Ed.). 2002. "Modern Organizations in Virtual Communities." Idea Group: Hershey, Pennsylvania.

OCLC. (2006). *OCLC Digital Archive Preservation Policy and Supporting Documentation*. Dublin, OH. Retrieved on 22nd March 2011 from http://www.oclc.org/support/documentation/digitalarchive/preservationpolicy.pdf.

Ozkan, S. and Karabacak, B.(2010). Collaborative risk method for information security management practices: A case context within Turkey. *International Journal of Information Management*, *30*(6), 567-572.

Patel, A., Qassim, O. and Wills, C. (2010). A survey of intrusion detection and prevention systems. *Information Management & Computer Security, 18*(4), 277-290

Pearson, D. and Webb, C. (2008). Defining file format obsolescence: A risky journey. *International Journal of Digital Curation, 3*(1), 89-106.

Porter, B. (2002). *802.11 Security*. O'Reilly: North Sebastopol, USA.

Rosenthal, D.S.H. (2010a). Keeping bits safe. *Communications of the ACM*, *53*(11), 47. doi: 10.1145/1839676.1839692.

Rosenthal, D.S.H. (2010b). Format obsolescence: Assessing the threat and the defenses. *Library Hi Tech*, *28*(2), 195-210. doi: 10.1108/07378831011047613.

Rothenberg, J. (1995). Ensuring the longevity of digital documents. *Scientific American, 95(*1), 24-29.

Rowlingson, R. and Winsborrow, R. (2006). A comparison of the payment card industry data security standard with ISO17799. *Computer Fraud & Security, 3,* 16-19.

Saeednia, S. (2000). How to maintain both privacy and authentication in digital libraries. *International Journal on Digital Libraries, 2*(4), 251-258.

Singh, S. (2003). *Digital library: Definition to implementation*. Ranganathan Research Centre: Delhi. Retrieved on 15th April 2011 from http://www.oocities.org/esukhdev/lecture_rcc.pdf.

Siponen, M. and Willison, R. (2009). Information security management standards: Problems and solutions. *Information & Management, 46*(5), 267-270.

Smedinghoff, T. (2004). *The state of information security law: A focus on the key legal trends*. Retrieved on 30th March 2011 from http://ssrn.com/abstract=1114246.

Sturges P., Davies E., Dearnley J., Illiffe U., Oppenheim C. and Hardy R. (2003), User privacy in the digital library environment: An investigation of policies and preparedness. *Library Management, 24*(1/2). doi: 10.1108/01435120310454502.

Sutlieff, L. and Chelin, J. (2010). 'An absolute prerequisite': The importance of user privacy and trust in maintaining academic freedom at the library. *Journal of Librarianship and Information Science, 42*(3), 163-177. doi: 10.1177/0961000610368916.

Thompson, S. (2006). Helping the hacker? Library information, security and social engineering. *Information Technology and Libraries, December,* 222-225.

Turle, M. (2008).Data security: Past, present and future. *Computers & Security, 25*, 51-58.

Whitman, M.E. (2003). Enemy at the gate: Threats to information security. *Communications of the ACM*, *46*(8), 91-95. doi: 10.1145/859670.859675.

Zimerman, M. (2010). Protect your library's computers. *New Library World, 111*(5/6), 203-212. doi: 10.1108/03074801011044070.