AN ALTERNATIVE APPROACH TO SECURITY CONCERNS IN DIGITAL LIBRARY INITIATIVES

N. NAGHSHINEH, Ph.D. Student
Faculty of Psychology and Educational Sciences
Tehran University, Tehran, I. R. of Iran
email: dialog@ncda.net

Abstract - Digital preservation refers to an important aspect that most Iranian digital library initiatives lack. This component is Security. While a digital library may be misconceived as an information supply space without a physical form, the same security concerns in force for operation of physical library are valid. Unlike the analog materials, all digital copies of digital materials are original. A digital library can provide a tool whereby the user can authenticate the same material without compromising sensitive watermark key data. Watermarks must be impervious to premeditated alterations and tampering. A national digital watermark assignment center is proposed for sharing the burden of development and distribution of such security features to content developers.


INTRODUCTION

Like any other society or social group, the Iranian librarianship follows a certain behavioral pattern deeply rooted within its history and social standing. It has its very own set of creeds, unwritten rules of engagements, truisms and fallacies. Many of the courses have not kept pace with social and economic changes, which pervaded a sense of obsolescence. It is as if the Iranian librarianship is stuck within time warp twenty years in past. This has led to a “bandwagon” effect. People try to climb aboard often without a foggiest notion where this bandwagon is not heading nor what is the tune. Many Digital Library initiatives in Iran have been victimized by this.

There is law in computer science known as “Ninety-Ninety Rule”. I like to modify this rule as follows: “90% of digital library planners in Iran are stymied by 10% of issues they ignore 90% of the times”. With the exception of libraries, containing digitized material catering to a select client group (mostly national security, intelligence or law enforcement); the planning should be based on a market-model. To put it another way, the success of any planning for a digital library hinges upon treating the patrons as clients and to touch upon issues such as brand loyalty, customer retention and niche market development. These are concepts alien to Iranian library planners in general. There is also an issue of transitions versus transformation. A digital library planning may be driven more by the
need to incorporate and develop assets and capabilities rather than transforming the
pre-existent social transactions prevalent in information gathering, storage and processing.
This would essentially lead to mere transplantation of usual library processes within digital
domain. The problem here is that given the general technology-intensive nature of the
digital library, two issues become of paramount importance: Stability and Security.

If we use product development process as an analogy for digital library planning, then
we will need to address the issue of life cycle costs. In government-dominated economies,
such as the Islamic Republic of Iran, life cycle costs are a luxury since most libraries are
propped up by direct government allocations. In a business approach, the life cycle cost
is determined by the size of the targeted customer’s base and the projected revenue (both
principle and residual) that, in turn, would determine the return on investment or ROI
profile. It is this profile that would determine both the scope and depth of upgrades. The
operating system that we had adopted for our digital library may not be around in 5, 10
or 15 years from now. Given the progress in scanning and digitization techniques, some
of our digital materials may become obsolete due to the force of technology. If we are a
museum digital library, this would mean that we might need to re-digitize our collection
(originally scanned from pictures at say 24-bit resolution) using an entirely different set
of standards or specs (say high-resolution 3D laser scan). In essence, we would be in a
process of making constant content adjustment with shifts in technology as well as customer
demands. This is a challenging area that could even have an impact on digital preservation
initiatives. Magnetic and optical storage media can deteriorate rapidly. Even when you
have applied Tempest standards for protecting your magnetic data, the magnetic damping
effect will deteriorate the content with frequent use. This would mean regular upgrades
of hardware and software as well as the content to ensure continuing availability of the
collection to the patrons.

Digital preservation points out to an important aspect that is amiss in most Iranian
digital library initiatives. A crucial component of the 10% that the planners fail to appreciate
90% of times is Security.

When speaking of security, our mind inadvertently drifts towards unauthorized intrusion,
hacking and cyber terrorism. While these considerations are valid for any network asset
building initiative, they just touch the tip of the iceberg.

SECURITY CONCEPTS

Security is of different meanings to different people. Many languages, and per se cultures,
maintain different shades of meaning for the word. In Persian, for instance, there are
several variations to the word that cover concepts from imperviousness to safety to
resistance. Perhaps Iranian history had something to do with it. A quick scan for the word
on dictionary.com provides the following meanings for “Secure”:

1. Free from danger or attack: a secure fortress.
2. Free from risk of loss; safe: Her papers were secure in the vault.
3. Free from the risk of being intercepted or listened to by unauthorized persons: Only one telephone line in the embassy was secure.
4. Free from fear, anxiety, or doubt.
5. Not likely to fail or give way; stable: a secure stepladder.
   Firmly fastened: a secure lock.
6. Reliable; dependable: secure investments.
7. Assured; certain: With three goals in the first period they had a secure victory, but somehow they lost.

The list goes on, but in the interest of brevity, I am confining my discussions to these seven explanations. While a digital library may be misconceived as an information supply space without a physical form, the same security concerns in force for operation of physical library are valid. Setting up a library in the real, physical world requires compliance with numerous building and operational codes. A librarian may not need to be aware of all of them, only those that would impact on the processes she/he is responsible for. Nevertheless, the security of a physical library is a function of its building, collections, staff, patron and hundreds of other details that make a library larger than the sum of its parts. The concept, here, is organizational security. Organizational security is a series of interrelated skills in threat assessment, threat reduction, physical security, operational security and disaster survival. Organization security implies the existence of a risk management plan to deal with contingencies.

Few Iranian libraries, if any, have an organization security protocols. Nevertheless, they are eagerly racing towards digitization. Some do this out of concern for their collection preservation; others may have diverse concerns such as compensating for shifting patron usage behavior, increased diversity of information media or efficient information delivery. Whatever the reason, their approach is manifested with vulnerability because of lack of an organization security protocol.

Outlining a Digital Library Security Plan (DLSP) does not entail hiring an ex-intelligence officer; however, putting all your eggs in a computer security analysis basket does not cut it. Some do tend to equate virtual library with digital library. However, this is just a play of the words, even something as ephemeral as human thought needs a physical brain to exist. A digital library may not be concentrated in any single location, but still it needs to be rooted within a physical location (whether a hard disc or a server frame) to be accessed. A DLSP is based on both the physical components as well as the processes that make up a digital library.

A DLSP is an exercise in identifying potential vulnerabilities, minimizing them or at
least make them less attractive to potential attackers. On the other hand you cannot attain a 100% security. What is often done is to carry a security audit to highlight these vulnerabilities. The audit findings should be kept confidential and secure. Such security audit is carried out at several levels, which includes staff members, information systems, physical premises and content distribution and access. The audit provides you with two lists:

1. List of threats and vulnerabilities
2. List of possible controls and countermeasures

These two lists would also provide you with means for carrying out a cost benefit analysis for the possible solutions. A digital library dealing with clinical toxicology and protocols for treating patients may require more robust anti-tampering systems than say a school digital library. Each area of vulnerability may be scored based on process criticality so as to implement proper trade-off between control and productivity.

If we go back to our earlier definition of security, and map it against the common concepts in library operations and processes, we could arrive at a rudimentary benchmark for setting up our risk management system and a basic DLSP.

*Free from Danger and Attack* could be translated into safe operational environment or protocols and having contingencies in place.

*Free from Risk of Loss* could be taken as implementation of preservation control. This may include implementation of say a digital vault concept (Such as those used by the banks for protection of their critical data) or building in the system the ability to reconstruct the lost resource. However, if the DLSP is to be devised in case of a digital library that is a manifestation of a non-digital collection, it should include preservation protocols for such material as well. For instance, in case of manuscripts, once digitized, they need to be sealed and stored in a secure storage area that may not be co-located with the physical library.

*Free from the Risk of being intercepted or listened to by unauthorized persons:* This is same as patron classification within a physical library as to manage access to collections. On the other hand, within an organization it means staff clearance scheme for access and manipulation of data. Within an automated library, this would require the patron to waive some of his/her rights to be allowed access. His/her usage information would be collected and stored automatically. In a sense, a patron surrenders a portion of his/her privacy to gain access to an information source or service. In this respect, this is no different from the instance when one applies for a credit card. However, one needs to provide assurances that the information thus obtained would not be available to third parties. If the access and distribution of information were being carried out electronically over open systems such as internet, this would entail implementation of secure communication algorithms.

*Free from Fear, anxiety and doubt:* In my opinion this is the most delicate aspect of
security planning for a digital library. In a physical library we often take it for granted that many patrons take the library as a secure environment whose collection serves as arbitrator of uncertainties. It is very important to translate this feature to digital domain where in a sense the burden of proof and credibility falls on the user. This may require more investment in GUI, interactive help desks and confidence-building representation of information. The content verification systems have been suggested, but these may prove too costly. Some have investigated mechanisms similar to those employed in e-commerce. Perhaps, developing a verification agent, or veribot, would be a possible solution. A veribot would be a secure software program that a patron could apply to the content received over the internet as to verify whether it is indeed a content produced, held and cleared for distribution by the given digital library. In fact, in an era when a digital document can exist in many places without the author’s knowledge, content management and digital rights management provides the single most daunting task in DLSP. The challenge, here, is finding that elusive balance between the need to provide trouble-free access to information as well as ensuring that the intellectual rights of the content creator is safeguarded.

DIGITAL CONTENT PROTECTION

Unlike the analog materials, all digital copies of digital materials are original. The technology for ripping digital material is so cheap that boggles the mind. Counterfeit DVDs in Iran sell for as little as under US$10. Some pirated titles are even tampered to play on just about any DVD players. However, DVD is not the only targets … expensive software licenses have given rise to a "Code-Breakers". Games such as Medal of Honor are now available in Persian not to mention a plethora of software titles worth thousands of dollars available at $2 a CD. Most of these digital materials come with embedded protections such as serial numbers, online activation requirements and such. Yet, they are cracked. What would happen if we set up a National Digital Library of Music and then find that our digital content, often available free of charge, are ripped and repurposed for financial gain by a third party? The issue becomes heightened when our collection deals with digital materials outsourced whose copying would precipitate in an infringement of copyright and perhaps less favorable licensing agreements. Because of the proneness of digital media to tampering, many Iranian courts do not consider them as evidence, unless you could provide proof beyond doubt of original ownership. Even when you catch a person who has made an unauthorized copy, you will have a hard time proving intentional malice.

For instance, many magazines and dailies in Iran make a heavy use of internet for features as well as photos and graphics. Few of them make reference to the original article or owner of the material. With exception of news agencies, practically none pays royalty. There are no newspaper syndicates in Iran and, therefore, there are no burden sharing for commissioning of neither syndicated columns nor content conversion for major news
entities such as CNN.

For a specialized digital library that needs the continued trust of its patrons in its content, it becomes imperative to protect its materials not only against theft but also to assure its user of its origin. But what is the best way to accomplish this?

One way is to adopt a security shell for delivery of digital content. The security shell is activated when a client connects to the digital library and monitors compliance with digital rights management protocols. One such system was adopted by companies such as netlibrary (www.netlibrary.com). In its early days to use the netlibrary, a client needed first download software to read the materials. The materials loaned electronically, were downloaded with a time stamp making the copy inaccessible after a certain time. To borrow further titles, a client needed to return or check-in the previous material. Later, Netlibrary opted for an online system, whereby the client was required to read the material online. A client was assigned a bookshelf where s/he could have stored his/her favorite titles. Even the annotations were made and stored online. Based on the copyright setup for individual scheme, a client could make a certain amount of copies. However, the system had a feature that monitored systematic copying. The netlibrary system is capable of displaying e-books in various e-book as well as PDF formats. Figure 1 shows a sample page of the system.

![Sample page for netlibrary.](image)

One should note, however, that such a security shell might in essence be in conflict with the right of privacy. Another system, used for document delivery, is ARIEL. ARIEL is a system developed by the Research Library Group (www.rlg.org) for electronic delivery of scanned documents. However, once scanned and delivered, there is no way to monitor further copying of the material. For visual materials, such as pictures, companies such as Corbis (www.corbis.com), employ a system whereby the pictures are displayed in low-resolution format with a trademark imprinted on them. The image has enough quality for the potential buyer to make a decision as to whether order the original print. But once the original print is obtained it could be recopied and distributed.

It, thus, becomes necessary to find a way to tag the materials made available digitally,
while this does not retract from the quality, to ensure the proof of ownership. Given the digital nature of the materials in question, such tagging is quite possible. It can provide the managers of digital libraries with the capability of identifying the route of copied material.

One way is to hide certain data indicating the origin of the material (collection, publisher, author, date of digitization...) within the digital material itself. This method of data hiding is known as steganography, a Greek term for "Covered Writing". Government legal tenders, such as paper money or treasury bills, have data hidden in them to prove their authenticity. Iranian passports, for instance, have embedded security (hidden data) features that are revealed under UV light. German company, Siemens, demonstrated a visa authentication system during the world summit on information society, whereby data such as the visa bearer's name, address, physical features and even picture is embedded within the visa.

Embedding a digital tag within a digital material is known as digital watermarking. Digital watermarks can be done in two ways: Perceptible where you can actually see the mark and imperceptible where the mark is hidden.

a) Now is the time for all men/women to ...  
   Now is the time for all men/women to ...

b) Now is the time for all men/women to ...  
   Now is the time for all men/women to ...

Figure 2: Imperceptible watermark embedding. Note that the letter "i" is slightly shifted to the left. (courtesy ACM and the Bell laboratories)

The most important stage is generation of the watermark. The watermark must contain unique information that would identify the owner. Some government digital libraries, such as those dealing with national security, have secured assets for developing such digital watermarks. However, the ability to develop unique, tamper proof watermarks is prohibitively expensive for most digital libraries; therefore, it is often suggested to outsource this service. There are companies that provide such service, not only by generating unique signatures and embedding algorithm but also by providing the software to read them. One such software is Digimarc Mediabridge. It can take a scanned or photographed image and, then, read the watermark embedded into it. If the watermark is generated by this software, it contains information that acts as a unique index. This index is looked up on the Digimarc server and points to a web link. This link is opened and appropriate information about the author or image can be displayed and necessary software can be loaded. This is done with
no extra involvement or information from the user, it is carried out automatically. The watermark is said to act as a ‘bridge’ between the image and the information, hence the name for the software. Currently, there are several software on the market that can help content managers to incorporate various forms of watermark within their digital assets. Many of these have been designed for protection of textual and image assets. But, there are a number of software that can be deployed for water marking of digital stream broadcasts, DVD, Digital Audio as well as CD-ROM. In fact, one software uses the digital watermark as an access control. In this system, the digital watermark is incorporated in the label. The CD-ROM contains a program that could identify such watermark. To access the CD-ROM, the user needs to run this program, scan the label containing the watermark, and upon authentication, would have the content of the CD-ROM decrypted on the fly for use.

![Image of Digimarc software screen](image_url)

**Figure3**: Sample of digimarc software screen. (courtesy syboldreports)

Of course, depending on the criticality of the information being distributed or made available to your client group, you may need to employ one or several watermarks that provide the best fit with your security concerns. The DLSP will provide you with some operation parameters. For instance, in case of legal documents the watermark could be embedded containing not only the collection data, but also data on transmitting and receiving terminals. In ARIEL example, for instance, the document delivered could contain hidden information on the order and the person who has ordered it.

One should be aware that the higher a watermark resists tampering, the higher would be the cost of its design and embedding. DLSP will indicate which portion of the digital collection needs overall imperviousness against manipulation and what portions do need to be resistant up to a point. The key is to strike a balance between active countermeasures and passive ones. For fraud prevention, it is understood that one would need a robust watermark. Nevertheless, it is also useful and less costly to use watermarks that resist tampering up to a certain point. These are also known as fragile watermarks. Fragile trademarks would stave off non-consented replication by all but the most die-hard.
One should note that when copying a digitally signed document, the perpetrator often employs methods such as lossy compression (image format conversion), cropping or re-pixelization. These methods could extract the watermark or render it useless. Once again, the issue of balancing security considerations against the unimpeded flow of information becomes paramount. There is also, as mentioned earlier, some privacy concerns. Personally, I believe that there is no basic incompatibility between watermarking and privacy. They cover two entirely different realms. Verisign Inc., a California-based internet security company, can provide concerned users with digital IDs for a nominal annual subscription fee of US$14.95. These digital IDs either could be incorporated into your email to ensure the recipients of authenticity or could be used to encrypt the content of the email as well as the attachments. A similar system could be envisioned in a DLSP where the users would receive digital IDs that would not only enable them to verify the watermark of the content supplied, but also would tag the data in a way that would identify the authorized (or unauthorized) replications committed. This could be implemented in compliance with e-trust website protocols.

There is also a possibility that some of your materials have been used in other publications (whether web-based or otherwise) and a reader wishes to authenticate the material. A digital library can provide a tool whereby the user can authenticate the same material without compromising sensitive watermark key data. The idea of a Veribot was developed to address this issue. However, presently there are no clear set of standards for development of such interfaces. However, e-commerce solutions such as online SSL authentication and encryption may be adapted for this purpose. Showtime Arabia, a broadcasting company based in the UAE, uses smart cards for decryption and delivery of its programming. The activation period is only for a year. However, it is quite possible to use the same technology against piracy. The signal could have an embedded watermark coupled with smart card unique data. Therefore, even if someone makes a copy of the program, it could still be traceable. Figure 4, demonstrates such a system earmarked for digital broadcast authentication and verification. This process is generically known as digital fingerprinting. In fact, digital fingerprinting is a technique of choice for tracking illegal copies. Systems such as those researched by NHK have a high initial capital investment, because several security algorithms are used for embedding within the watermark. The watermarks would be unique for every digital material transmitted with security robustness that could even survive the digital-to-analog ripping and digital repurposing (Figure 5).
At any rate whatever anti-tampering or anti-piracy protocol your DLSP suggests, it must fulfill the following qualifications when deployed within your system:

**Transparency:** The presence of the watermark should not have any detectible effect on the delivered quality, whether an image, video, audio or text survivability: the watermark data must be able to endure and survive when subjected to a broad range of deformation and misrepresentation included, but not limited to internet transport distortion, compression and re-pixelization.

Data throughput or potency: The watermark technique must lend itself to high throughput data transmission, etc. as it must be able to extract the data from a portion of the medium. In a way, this concept is similar to that of an early hologram plate. A hologram plate contains not image data but patterns of interference. Thus, even if you break it, you could reconstruct portion of the image from the shards.

**Efficiency:** Encoding, and decoding of watermark data must be relatively cheap, and least resource intensive (processing and computing).

Security: Watermarks must be impervious to premeditated alterations and tampering.

**CONCLUSIONS**

In countries such as Iran that have a disparate ICT infrastructure, the digital library initiatives could quickly find themselves quagmired in either inadequate security policies or the never-ending spiral of upgrades. Because of the prohibitive cost of incorporating
an embedded security, dependence on government budget allocation as well as dearth of competent computer security firms, the measures for content security and digital rights management often fall short of being efficient. Perhaps in such cases it would be advisable to have the digital library stakeholders establish a security consortium to share the burden of the cost. One other alternative is to secure government funding for establishment of a reference digital security asset that would issue unique watermarks to those intent on distributing their content over the network or through digital format.

Another suggestion is to portion the cost of protection of digital assets based on their criticality. Cost-benefit studies could identify these assets as well as providing a cost projection for appropriate protection measures.

At any rate, it is imperative to have a robust authenticating protocol in place for tagging digital assets in order to maintain the credibility of the resource in the minds of the patron. The issue here is to transfer the credibility precepts taken for granted within a physical library setting, into a virtual space. Old library transactions and trust-building models may not make the transition unscathed. Therefore, it may become necessary to adopt e-commerce models in trust building.

FURTHER READING (E-BOOKS)


Digital Watermarking Software:

Photo Water Mark Professional 4.4 (http://www.unidreamtech.com)

Photo Watermark Professional is a powerful, digital-photo-watermarking tool that combines simplicity and efficiency and simplifies the process of creating and applying watermarks to multiple files. It uses a mixture of text and graphics, drawing watermarks of any complexity. The automatic objects can extract various EXIF data from your images. The multilevel transparency settings allow you to choose how your watermark will affect your pictures. The program supports JPEG, TIFF (including multi-image TIFF), BMP, GIF (including animated GIF), PNG, and JPEG 2000 files. The built-in watermark manager helps you manage multiple watermark files and easily switch. Its batch features include combinations of watermarking, converting, resizing, renaming, cropping, rotating, fine-tuning and printing. A built-in FTP client lets you upload and download photos directly to and from your server. [this is a company-supplied description]

Digimarc Media Bridge (http://www.digimarc.com)

Digimarc MediaBridge™ allows Digimarc-enabled brochures, direct mail pieces or packaging to link to special online offers, product information or buying opportunities. Whether triggered by a reading device within an in-store kiosk or by a PC-tethered Web camera, MediaBridge offers marketeers an innovative, new way to Integrate print and electronic campaigns or deliver valuable content to consumers. [This is a vendor-supplied information]

Stirmark Benchmark 4 (http://www.peticolas.net/fabien/software/stirmarkbenchmark_4_0_129.zip)

This is a generic research tool aimed at designing and developing a fully automated and complete benchmark suite for digital watermarking schemes. Its other aim is to make this benchmark one of the reference services for researchers, suppliers of watermarking technologies, industrial users and end users. The software is free to download as long as it is employed for research purposes.

First launched in 1999, VeriData iDem software was specifically developed in conjunction with forensic scientists and crime-scene investigators to overcome the problem of digital image integrity. It can validate images created with almost all types of digital cameras and can be used with a wide range of forensic imaging, fingerprint enhancement or document examination systems.

GENERAL REFERENCES


