European Digital Libraries: Web Security Vulnerabilities

Abstract

Purpose – The purpose of this paper is to investigate the Web vulnerability challenges at European library Web sites and how these issues can affect the data protection of their patrons.

Design/methodology/approach – A Web vulnerability testing tool was used to analyze 80 European library sites in four countries to determine how many security vulnerabilities each had and what were the most common types of problems.

Findings – Analysis results from surveying the libraries show the majority have serious security flaws in their Web applications. The research shows that despite country-specific laws mandating secure sites, system librarians have not implemented appropriate measures to secure their online information systems.

Practical implications – The findings serve to remind librarians of the complexity in providing a secure online environment for their patrons and that a disregard or lack of awareness of securing systems could lead to serious vulnerabilities of the patrons personal data and systems. Lack of consumer trust may result in a decreased use of online commerce and have serious repercussions for the municipal libraries. Several concrete examples of methods to improve security are provided.

Research limitations/implications – Further research on library vulnerability throughout the world can be taken to educate librarians in other countries of the serious nature of protecting their systems.

Originality/value – This paper serves as a current study on data security issues at Western European municipal library Web sites. It serves as a useful summary regarding technical and managerial measures librarians can take to mitigate inadequacies in their security implementation.

Keywords Digital libraries, European libraries, Data security, Web vulnerability, UK, France, Spain, Germany

Paper type Research paper

Introduction

There has been a tremendous growth in the use of digital public libraries, and consumers have benefited from the growth and ease of services, such as the ability to conduct online browsing and research from the comfort of their homes. Patrons have recently grown accustomed to the increase in digital library services and the ease at which they can access global information that was once extremely difficult to retrieve from traditional physical libraries. For example, although the use of e-books initially got off to a slow start, their use has increased in libraries and will continue to grow due to a broadening variety of access formats and available functions, such as full-text searches (Gernand, 2006). Gernand also mentions that electronic journals and robust research methods are becoming widely popular.

However, along with these advantages come concerns with the technology related to online services. Data protection, consumer trust and securing personal computers from malware have become issues with patrons and librarians. Fox (2006) explains one problem with online commerce is that most users and librarians are not aware of the security issues affecting their systems and networks. He states that one part of providing better overall library security is to educate staff and patrons on concerns related to information security threats.
Internet based applications, such as digital libraries, require appropriate security mechanisms because millions of participants across the globe access these sites to obtain services. Tighter security is one of the factors that may increase the value and utility of these applications and can contribute to higher levels of consumer trust with using online services (Chen, et al, 2006). When vulnerabilities are exposed and make headlines, user trust and confidence in online services is eroded. If these occurrences become commonplace in the library community, efforts to collect and preserve digital content could be hampered as patrons may decide not to use online content (Fox, 2006). Therefore, it is imperative that to provide a safe browsing experience, libraries take appropriate steps to secure the safety of their, and their patron’s, systems.

Purpose and Methods Used
The research was aimed at examining the following issues related to security of European digital libraries:

- What level of security exists among European library Web sites?
- What are the most common vulnerabilities?
- Is there a marked difference among countries in providing secure library systems?

The research starts with a survey of the literature to review the growth and market of the digital library marketplace. A discussion of legal requirements for securing sites is reviewed, as well as technical issues related to Web commerce and current security problems with libraries. This portion shows the current lack of literature and studies in the field of library security, especially related to Web security of European sites.

A methodology for testing Web vulnerability of 80 library sites in four countries is reviewed, along with an interpretation of results. Finally, implications of this study to the library industry along with recommendations are presented. Although this study is mainly aimed at librarians and systems administrators, the results of this research go far beyond Web vulnerabilities within library sites. The same vulnerability testing methodology could be used effectively by administrators and site owners of all online commercial sites in order to provide better protection of their customers. Also, library patrons may find the results relevant when determining the level of trust in placing with libraries, and contacting librarians to ensure that secure systems are in place to protect their private data.

Framework of the Library Market and Security

Library Marketplace
Vaugh (2005) states that more and more information is not only available online, but preferred online, thus leading to a growth in service providers such as digital libraries. These entities are providing a wider variety of services, and will continue to do so as Internet technologies expand. Chowdbury (et al, 2006) claims that public libraries will need to shift from solely providing access to knowledge to acting as a platform for storing and disseminating local community knowledge within a global context by using new technologies, such as Web 2.0.

Digital libraries can enable world-wide access to an ever-growing quantity of distributed information and knowledge. This is especially true where consortiums and groups of libraries are working together in collaborative management, electronic publishing and document delivery (Raitt, 2000). The authors further indicate that European libraries are undertaking a number of initiatives at the local, national and international level in order to improve patrons’ access to digital information, thus attracting a wider audience to their services. For example,
The European Library initiative comprises 48 European national libraries that have a single portal for world-wide audiences to access collections from national libraries (The European Library, 2009).

Although no official government sites list all public digital libraries throughout the world, several directories of these libraries have been compiled by various sources. A register of UK digital public libraries has a total of 219, with the majority containing at least a minimum service level of catalog listings (Harden & Harden, 2008). A Libweb catalog of a select number of Western European digital public libraries lists 93 for France, 80 for Italy and 133 for Germany (Dowling, 2009, which shows a sizable market of digital libraries for Western European nations. The numbers for Eastern Europe, although growing, still represent only a fraction of that of the West, with number on Libweb showing 10 for Bulgaria, 12 for Romania, five for Serbia and 15 for Russia (Dowling, 2009). With the increasing number of libraries that provide digital content, the number of patrons who access these services will continue to grow. However, positive growth does not come without issues that librarians need to address in order to protect their patrons, such as meeting legal obligations for security and providing safe technology.

Legal Aspects
As consumers are flocking to online service providers, many of these same people are becoming increasingly concerned about the safety aspect of using these sites. A plethora of news articles about hacking attacks and loss of personal private information contributes to a potential loss of consumer trust. Because firms have often been unable to self-regulate their industries in providing secure environments, governments have enacted legislation to provide a minimal level for secure systems and provide better data protection of consumer’s personal information. Individual European countries have laws that mandate data protection of personal data, and state that online service providers must maintain secure computer systems through a variety of technical and procedural processes.

In the UK, the Data Protection Act of 1998 specifies the legal obligations that companies have in protecting personal data and governs technical responsibilities for storing data. Other legislation introduced in 2008 enhanced the Act to include notification requirements for breaches as well as giving government entities the power to conduct security audits. (Castro-Edwards, 2008).

The French Data Protection Directive (Law Nr. 2004-801) came into force in 2004 and relates to protection processing of personal data as well as sanctions for breaching the law (PRIVIREAL, 2005). Italy’s Data Protection Code - Legislative Decree No. 196 of June 2003 requires entities to implement specific technical, logical and organizational minimum security measures, but does not require notification of breaches. (Linklaters, 2009a).

Spain’s Data Protection Directive was entered into force in April 2008 and imposes specific security obligations that must be used to protect personal data, as well as the obligations that entities must take to notify consumers when breaches occur (Linklaters, 2009b). Legal sanctions have been imposed for Spanish firms breaching security and data protection laws. In June 2007, Spain’s Supreme Court imposed a one million Euro fine against a firm for breaching their Data Protection Act and not complying with computer security regulations (Privacy & Data Protection, 2009).

Security Problems and Breaches
Library administrators have unique security concerns compared to other industries. They must provide secure systems to employees as well as the general public who use computers in the libraries, and also patrons visiting their Web site (Balas, 2005). She suggests that systems librarians are often unprepared for managing networks and newer technology, and they
should take steps to be more well-informed on security topics. In addition to protecting their own systems, librarians must take precautions to ensure that online services do not jeopardize their patron’s privacy and personal information (Breeding, 2005). He indicates that librarians are developing personalized services with online e-commerce transactions and creating accounts that include personal data and the ability to exchange money, so it is important to implement safeguard technologies. Librarians often find themselves attempting to secure several functions related to online service. According to Fox (2006), librarians need to preserve and guard the digital content, some of which can be very valuable such as information about patents or politically sensitive research. They also need to protect the private data and confidentiality of their patrons.

In order to access many library services, patrons must establish an account using personal information such as address or telephone number. Libraries may incorporate the ability to make payments with credit card numbers, another delicate piece of data that needs to be secured. Thompson (2006) states that because subscriptions to some proprietary library databases are expensive, hackers may find these resources of particular interest. Thus, there is a variety of confidential personal and research data that make libraries targets for malicious security attacks and penetrations.

Specific research on security breaches at libraries is sparse, especially related to European digital libraries, although more literature does exist on North American library attacks. In summer 2002 and May 2004, there were two security breaches at the Indiana University in the US. Hackers had managed to find loopholes in security, and it took significant time and effort to restore and upgrade the system to include new security defenses (Cheng, 2005). Another attack on a library occurred at the University of Notre Dame digital library where service was compromised because of non-encrypted authentication on Web content services (Fox, 2006). Currently, no Web application security studies have been completed specifically involving Western European sites, therefore showing a lack of literature in this field.

**Web Application Security**

This research is aimed at a cross-section of staff librarians and systems administrators who may not have in-depth application security knowledge. However, it is still important for most library employees to have an overall understanding of some of the security issues related to Web applications in order to better understand how to protect their systems. Some types of security issues are more common than others, and it is valuable to understand these in order to install and maintain specific methods for vulnerability protection. Web security firm Cenzic (2009), indicates that during 2008, almost 80 percent of Web-related flaws were caused by Web application vulnerabilities with the three most common types being: a) Cross-site Scripting (XSS)), b) Denial of Service and c) SQL injection. Fox (2006) and the SANS™ Institute (2009), a worldwide security organization, the also mention cross-site scripting as the most common Web application vulnerability which can cause a variety of vulnerability holes and hacking attacks.

Another universal problem with Web application security is the lack of updating software versions or neglecting to install security updates. Security experts recommend that prompt upgrades of software and patches be applied, but according to Cox (2002), there are several reasons that Web administrators may not do so including: a) developers install the default software and forget it needing updating, b) lack of consideration of security flaws and c) lack of upgrading software correctly. A third factor in weak Web sites is lack of effective coding practice during design and development phases. Viega & McGraw (2001, p. 5) states that many vulnerabilities are found in source code, and that careful thought should be given to security before coding begins (p. 24).
Methodology
The research in this paper was accomplished through analyzing 80 Western European digital libraries to determine the most common security vulnerabilities. The project consisted of three phases:

- Choosing a testing tool.
- Choosing a list of countries and sites to test.
- Running the analysis and analyzing results.

Choosing a testing tool
The first phase of this study was choosing a testing tool to scan for Web vulnerabilities. Several criteria were used to select the tool including cost, functionality scans and ability to run on a personal computer (PC). First, because the researcher’s PC would be used for this testing, it was imperative that the product could run on a PC using Microsoft XP operating system. Second, due to budget constraints, the software had to be affordable, costing under $100. Finally, the software had to be able to test for a variety of Web vulnerabilities, including cross-site scripting and outdated versions of Web software.

One software product did meet these requirements, N-Stalker Web Application Scanner 2009 Free Edition. This product was free and could report on a variety of Web application vulnerabilities including:
1. Cross-site scripting
2. Web signature attacks on IIS, FrontPage, PHP, ASP and others
3. Backup security checks
4. Old software checks (N-Stalker, 2009)

Choosing a list of countries and sites to test
The second phase of this project was to select four Western European countries where digital libraries are conventional, these were: a) UK, b) France, c) Italy and d) Germany. The Libweb global library catalog showed a significant number of digital Western European libraries, while very few were found in some Eastern European countries, such as only five library sites in Serbia (Dowling, 2009). Therefore, in order to gain enough sites for a valid statistical analysis, 20 library sites were chosen from the four Western European countries. The top 20 library sites were chosen from each list.

Each site had to be reviewed individually to ensure that the Web link on the Libweb listing did work. In one case, a link for a library in Strasbourg, France did not work, so another library site was substituted.

Running the analysis
After N-Stalker was downloaded and installed, an individual Uniform Resource Locator (URL) site name for an individual library Web site was inserted into the Scan Wizard box. This vulnerability scan was started and the software analyzed the specific site for a variety of vulnerabilities. On average, a normal test scan took 15 minutes, although the maximum time was two hours. All 80 sites were tested over a two-week period.

N-Stalker produced a technical report showing all vulnerabilities and divided them into three categories: a) high priority, b) medium level and c) informational issues.

- High level – critical vulnerabilities which could lead to high risk of damage or attacks. These issues should take precedence when implementing security designs and changes.
- Medium level – moderate ranked problems that could pose some risk to Web applications. These should be corrected after high-risk vulnerabilities have been corrected.
Informational – issues that probably pose little risk, but still should be analyzed by security developers in case any could lead to damage.

For each level, different types of vulnerabilities could be found and a number could be found for each. For example, the specific error ‘old apache version’ would be listed under the ‘high’ level categories. This error might be only found in once in a Web site scan, or perhaps could be found in a high number of occurrences for the library. The results were tabulated into an Excel spreadsheet for analysis.

Results Analysis

Table 1 shows overall testing results for each of the four countries: a) UK, b) France, c) Italy and d) Germany. The first column shows the three levels of vulnerabilities: high, medium and informational. Each level is then divided into separate rows listing various results for each vulnerability level: including: a) the total number of errors within the level for a country, b) the number of different types of issues, c) the number of sites for each country that had problems for that level, d) the range of issues for that level and e) the average number of vulnerability results for that country.

The UK showed the greatest number of both high-level (197 total, 16 types) and medium-level (7768 total, 302 types) vulnerability errors. This was significantly greater than the total number and different types of errors for libraries in the other countries. France and Germany displayed similar numbers of high and medium levels issues, while libraries in Italy displayed 10 high-level errors and 41 medium. The number of informational messages for all countries was close, with the average ranging from 5.4 (UK) to 6.7 (Italy).

Table 1. Vulnerability Testing Results

<table>
<thead>
<tr>
<th>Level</th>
<th>UK</th>
<th>France</th>
<th>Italy</th>
<th>Germ.</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>197</td>
<td>23</td>
<td>10</td>
<td>27</td>
<td>257</td>
</tr>
<tr>
<td>Types</td>
<td>16</td>
<td>5</td>
<td>4</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>Site w/</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>Range</td>
<td>0-8</td>
<td>0-4</td>
<td>0-4</td>
<td>0-5</td>
<td>0-8</td>
</tr>
<tr>
<td>Average #</td>
<td>49.2</td>
<td>3.8</td>
<td>2.5</td>
<td>3.9</td>
<td>12.2</td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7768</td>
<td>161</td>
<td>41</td>
<td>215</td>
<td>8185</td>
</tr>
<tr>
<td>Types</td>
<td>302</td>
<td>11</td>
<td>7</td>
<td>50</td>
<td>315</td>
</tr>
<tr>
<td>Site w/</td>
<td>12</td>
<td>11</td>
<td>6</td>
<td>13</td>
<td>42</td>
</tr>
<tr>
<td>Range</td>
<td>0-237</td>
<td>0-3</td>
<td>0-3</td>
<td>0-49</td>
<td>0-237</td>
</tr>
<tr>
<td>Average #</td>
<td>647</td>
<td>14.6</td>
<td>6.8</td>
<td>16.5</td>
<td>195</td>
</tr>
<tr>
<td>Inform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>126</td>
<td>120</td>
<td>122</td>
<td>438</td>
</tr>
<tr>
<td>Types</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Site w/no</td>
<td>13</td>
<td>19</td>
<td>18</td>
<td>20</td>
<td>70</td>
</tr>
<tr>
<td>Range</td>
<td>0-2</td>
<td>0-2</td>
<td>0-2</td>
<td>0-2</td>
<td>0-2</td>
</tr>
<tr>
<td>Average #</td>
<td>5.4</td>
<td>6.6</td>
<td>6.7</td>
<td>6.1</td>
<td>6.3</td>
</tr>
</tbody>
</table>
Most sites contained some occurrences of either high or medium-level vulnerability issues. In the high-level category, seven German sites had vulnerabilities French sites had six and Italian and UK sites had four. The medium-level category had a greater number of issues, with 13 Germany sites having issues, 12 for the UK, 11 for France and 6 for Italy. Overall, six UK sites showed no occurrences of medium or high-level results, only one did not possess some result within each of the three levels. For French libraries, seven were free of vulnerability in both the high and medium-level categories, but all had at least some informational warnings. Italy had nine sites with no serious problems, and six German sites had this result.

The ‘range’ category shows how many different vulnerability types occurred for each site. For the high-level category, there were sites with no errors, but the UK sites did have a maximum of eight different errors, while France and Italy had four and Germany had five. Numbers in the medium-level category showed a significant range of issues, with the UK showing a maximum of 237 vulnerabilities and Germany containing up to 49 issues per site.

The average number of problems in the high-category ranged from 49.2 problems for UK sites to a low of 2.5 for Italian sites. In the medium-level category, UK sites had an average of 647 medium-level vulnerability problems, while other countries had much lower numbers.

Table 2 shows the vulnerability types within four common categories. The most common types of problems are cross-site scripting issues, with 6465 total scripting errors found in 474 pages. It should be noted that a single error could be found multiple times within a site. Another common safety concern is when older versions of software have not been updated to include new patches and fixes, with these results in this study showing 65 occurrences in 59 sites. There were 309 coding problems found in 18 sites. Finally, other miscellaneous types of issues, such as backup file issues or specific software problems, were found in 202 sites with a total number of 2157.

Table 2: Common Error Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Pages with Errors</th>
<th>Total Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scripting</td>
<td>474</td>
<td>6465</td>
</tr>
<tr>
<td>Old versions</td>
<td>59</td>
<td>65</td>
</tr>
<tr>
<td>Coding</td>
<td>18</td>
<td>309</td>
</tr>
<tr>
<td>Misc</td>
<td>202</td>
<td>2157</td>
</tr>
</tbody>
</table>

Implications and Recommendations

Evaluation results show that the majority of sites (65 percent) had either medium or high-level security vulnerability problems. Although it appears that UK sites initially contain a higher propensity towards security violations, it should be noted that two of the sites each had over 200 individual error results, which skewed results significantly within the medium-level category. Thus, if the results of these two UK sites were deleted, overall statistics for all countries would be much closer and may not show much of a statistical difference. However, even with this factored into the statistics, there was still a sizable number of different vulnerability types that showed up for most sites. This shows a serious issue with the security of all digital European libraries for all countries, which could result in both the possibility of putting their patron’s data and systems at risk as well as the prospect of losing consumer trust in their services.

According to Balas (2005), system librarians bear the brunt of securing library systems, but all librarians should be well-informed of the topic and issues in order to better protect the computer system and help patrons. It is imperative that library administrators keep up-to-date
on security releases and information provided by industry groups and vendors. According to Balas, (2005) groups such as CERT (Computer Emergency Readiness Team) and Software Engineering Institute (SEI) of Carnegie Mellon University are computer security organizations which provide updates to the latest security threats and ways to mitigate them. Hardware and software vendors such as Microsoft also provide security information that may be vital to library administrators keeping their systems secure.

Fox (2006) states that system librarians should check software against vulnerability alerts from CERT, and immediately install software patches and update their software to defend against attacks. Administrators can set up their systems to automatically download and install specific critical patches. This process has been successfully managed at the University of Las Vegas, Nevada, a community and academic digital library with a strong sense of technology in managing their resources and mitigating against security problems by a combination of securing technology, strong leadership and periodic training of library staff of the importance of security (Vaughn, 2005).

Digital library security should be thought of as a multi-dimensional approach with a combination of both technical and managerial solutions. While a variety of technical approaches have been discussed and can serve as one method to mitigate vulnerabilities, librarians should also consider the managerial concerns dealing with online security. Firstly, although firms must consider proper security as a business return on investment, cost and risk implications need to be analyzed when securing and implementing resources. With limited technology budgets, entities must have an economic rationalization when determining how much to spend on security. A 2008 survey of executives in large firms found that 53 percent said that their organizations allocated five percent or less of their overall Information Technology budget to information security (Richardson, 2008). He further explains that managers need to have an economic justification for their budgets. Mercuri (2003) states that firms should perform a risk analysis of their systems and then classify risks in terms of high, medium and low probability. Those vulnerability factors which have a high risk and can lead to severe repercussions for the firm should receive greater attention and funding compared to functions which have low risk of attack fewer consequences of impacting business functions.

Secondly, librarians should consider performing periodic auditing and logging audits of the Web application to determine if the system is vulnerable and to ensure higher levels of protection (Lampson, 2004). Not only is this a recommendation, but it is the law in several countries. In the UK, compliance audits are part of the Data Protection Act, which does include digital libraries (France, 2001).

Breeding (2005) indicates security threats against digital libraries will continue to rise and it is vital librarians take step to ensure that patron’s information is secure. He suggests that libraries that have basic Web sites and catalogue functions may not need enhanced security features. However, many libraries are now adding more electronic commerce features to their sites, such as entering personal data, passwords, reserving books and paying fines and fees. Thus, it is imperative that more secure systems are implemented as digital libraries increase their services. Huwe (2005) states that most security now involves a reactive rather than proactive approach, but with security threats associated with Internet technology, it is vital that libraries take a more proactive approach to securing systems and patron’s information. A proactive stance can protect the library’s online collection and Web services, as well as the needs of the library’s patrons who have difficulties in keeping track of their own systems security problems.

Security of Web sites is a major concern not only of the library industry, but also other industries and governments. This research could be further expanded to analyze schools, government and business sites in Western Europe and throughout the world. With regards to the libraries analyzed in this paper, results could be shared among individual libraries and
associations to instruct them on security issues. A follow-up analysis could be done in the future to determine what changes may have been implemented on the individual sites. Also, it may be interesting for additional investigation to determine if there are explicit factors that contribute to specific vulnerabilities among libraries compared to sites from other industries. This expanded research could contribute to understanding how to better protect entities against common vulnerabilities to their industry. Finally, it should be noted that although these results do show serious issues with library Web security, there should be a discussion on how some limitations could affect the results, and should be taken into consideration in upcoming analysis with this topic. First, the free edition of the N-Stalker software used in this testing inspects up to 100 pages in a specific Web site (N-Stalker, 2009). Because of this, it may be possible that the number and types of vulnerabilities for each library site could be higher. Further testing could be done with the full-feature version to determine if this is the case. Second, results could be skewed because of the amount of interactivity within the sites. Those pages with more interactive features could possibly contain more vulnerabilities. Further analysis could determine whether there is a correlation between high interactivity compared to the number of security susceptibilities.

Conclusion

The paper argues that unsecure digital library systems can lead to adverse consequences for both the libraries as well as patrons. There are several findings to this research that are relevant for library staff and patrons. First, Web vulnerabilities have been found to be common in the online environment and the majority of Western European libraries have critical (25 percent) or medium-level security problems (40 percent), leading to unsecured online commerce. Second, there are specific categories of concerns that are typically common in Web security. These could lead to attacks on the system, penetration of private data and a loss of trust by patrons.

Providing a safe online experience is a daunting task for today’s library staff because of the growing threats of Web vulnerabilities and constant hacking attempts. However, using a multi-dimensional approach to security, staff could mitigate problems. Managerial processes such as performing periodic audits and risk assessments can help determine which systems are most vulnerable and concentrate on protecting those functions, as well as inspecting the systems to analyze potential problems. Technical approaches are the second method of diminishing these factors. These include ensuring updates and patches are promptly installed and having coders and designers implement safe coding practices. These processes and procedures can be helpful for other libraries besides those in Western Europe, and can enhance their systems security and protect patron’s data.

References


