JLIS.it 10, 2 (May 2019)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12545

JLIS.it

# The progress of web security level related to European open access LIS repositories between 2016 and 2018

**Matus Formanek[a], Vladimir Filip[b], Erika Sustekova[c]**

a) University of Zilina, https://orcid.org/0000-0002-0611-1794
b) University of Zilina
c) University of Zilina

**Contact:** Matus Formanek, matus.formanek@fhv.uniza.sk; Vladimir Filip, vladimir.filip@fhv.uniza.sk; Erika Sustekova, erika.sustekova@fhv.uniza.sk
**Received:** 23 January 2019; **Accepted:** 20 March 2019; **First Published:** 15 May 2019

## ABSTRACT

This article focuses on the development of European institutional repositories web security in the field of Library and Information Science (LIS). Since the first analysis in June 2016, we have been still using the same three independent online tools to measure the web security score of these repositories. In case of data transfer, the qualitative aspects of the secured HTTPS protocol are as important as implementing the protocol itself. Our analysis is directly related to the previously published article where we examined the selected group of LIS repositories. Now, we are focusing on the summarization of improvements made between 2016-2018. These are based on periodical annual measurements. These may contribute to increasing the security level of repositories not only in Europe, but also with respect to the GDPR (General Data Protection Regulation) regulation which came into force a few months ago. It is significant because the GDPR focuses, inter alia, on the ways of sensitive data transfer over the internet networks.

## KEYWORDS

Web security; Institutional repositories; Digital libraries; Library and Information Science; Vulnerabilities; Web server.

## CITATION

Formanek, M., Filip, V., Sustekova, E. "The progress of web security level related to European open access LIS repositories between 2016 and 2018." *JLIS.it* 10, 2 (May 2019): 107-115. DOI: 10.4403/jlis.it-12545.

JLIS.it 10, 2 (May 2019)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12545



## Introduction

The security of computer online systems such as digital repositories is a widely discussed topic. The online environment is an ideal space for various cyber-attacks or the theft of virtual identities and other secret data. Nowadays, new issues are constantly being raised in the context of research and development of digital repositories and cyber security: the security and integrity of stored data/digital objects, metadata, personal data, credentials, and so on. Numerous articles have already dealt with this topic. Moreover, many researchers and security experts have recommended to encrypt sensitive data in any case: "sensitive data should be encrypted at all times, including in transit and at rest. No exceptions" (Kalman 2018). "Today, there is no such thing as non-sensitive web traffic, and public services should not depend on the benevolence of network operators" (The HTTPS-Only Standard, n.d.).

The sensitive data (passwords and logins, personal e-mail addresses… etc.) are stored in numerous online systems with a database, as well as in digital repositories. Therefore, we focus on security aspects. Our paper is organized as follows: introduction, which contains a set of information retrieval results together with a methodology description applied within further sections of this study.

Furthermore, the study consists of two sections. The first one deals with the increasing number of several digital objects available through LIS repositories from Europe. The second one deals with the evaluation of web security measurements within a three-year period.

The main goal of this paper is to collect and to summarize appropriate results, to analyse and to postulate conclusions concerned with web security of repositories related to Librarian and Information Science in EU. The paper brings an update of the previously published research results – Formanek&Zaborsky (2017).

To clarify the basic concepts and facts that we use in our analysis is considered to be a partial aim; however, a detailed description of technical issues could be omitted.

## State of the Art

The encrypted communication between the client and the server (e.g. digital repository) is considered to be the basic technology within web environment and many authors accept that fact. That approach prevents the disaffection of sensitive data during their transfer via HTTPS protocol as postulated by Kalman (2018): "Use HTTPS with a proper certificate and PFS (Perfect Forward Secrecy). Do not accept anything over non-HTTPS connections."

Whenever a user logs in with the web form their password is exposed in plain text on the network. This is a very serious security risk since network traffic monitoring is very common, especially at universities, as it is stated in DSpace manual (DuraSpace 2018, 63). Administrators should consider using HTTPS.

"The HTTPS protocol is widely used for a data transfer from a web server to a browser. Thanks to that we can browse all web pages. The main problem of the basic protocol is that HTTP (without "S" at the end) connection is not encrypted so it can be intercepted by third parties to gather data being passed between the two systems" (Entrepreneur Europe, n.d.).

JLIS.it 10, 2 (May 2019)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12545



As a result of that, the non-HTTPS communication might be intercepted, while a potential attacker is able to open and to analyse transmitted packets and discover login data concerned with the administrator or users and the entire system security is in danger.
As it is stated within the study by Adnan *et al.* (2017): "Compromised user credentials are a legitimate user account that has been taken over by a criminal or attacker… It is very effective for cyber-criminals to gain trusted relationships with the account owners. Such a compromised user's credentials ultimately result in damage incurred by the attacker at large-scale. Moreover, the detection of compromised legitimate user activities is crucial in competitive and sensitive organizations because wrong data is more difficult to clean from the database."
On the other hand, the HTTPS (with suffix "S") was designed by layering HTTP over Transport Layer Security (TLS) to provide end-to-end protection against network attacks (e.g. eavesdropping, man-in-the-middle attacks, etc.). The SSL/TLS certificates used in that technology enable increasing credibility of a target system and helping visitors with an easier identity verification related to the actual website.
The GDPR regulation deals with sensitive data protection during the transfer and storage processes. The personal data should be encrypted in every exposed digital environment.
When considering the paper context, our investigation activities related to HTTPS protocol implementation are concerned with a group of selected institutional repositories in Europe. However, that aspect seems to a problem round the world; and it is not concerned with Europe only. Our mission statement is to investigate only that area; and we have dealt with those issues for a longer time. We published a related paper to this topic in 2017 (Formanek & Zaborsky 2017). Roy Tennant (2017), the reviewer, postulated his attitude to that article as follows: "it is a wake-up call for all institutional repository managers to take a hard look at the security of their systems." Since that time, the problems closely related to security mechanisms within Librarian and Information Science seem to be worth of continual investigation for us. More information might be found within further sections of that paper.

## Methodology

In a direct response to the previously published article (Formanek & Zaborsky 2017), we outline the methodology as follows:
"We chose repositories by selecting them from OpenDOAR.org – the authoritative directory of registered academic open-access repositories. We used the following selection criteria:

- they should be institutional repositories,
- part of the content of repositories as well as their interfaces must be available also in English,
- the repositories should be located in Europe,
- their focus should include the LIS area (Library and Information Science)" (Formanek & Zaborsky 2017).

This registry listed about 33-34 digital repositories between the years 2016 and 2018; they were always sorted descending by number of available objects. Some repositories started up during this period. On the other hand, some systems disappeared or weren't functional. Because of this fact, the number of tested repositories was being changed every year.

JLIS.it 10, 2 (May 2019)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12545

Next, we studied home page/login page of all these repositories in more details. We tested the pages with three independent online tools as it will be shown later. The results were highlighted in the tables from which the score (the progress of security level) between 2016-2018 was evident. In this paper, we wanted to find out if GDPR had an impact on the security level of the selected group of digital repositories during last year (2018).

## Results and Discussion

### Section 1

At first, we found the current number of available digital objects in OpenDoar's records for each repository. Subsequently, we summarized and recorded appropriate results, while the data was recorded in June 2016, 2017 and 2018. After having completed the data evaluation, the results are visualized (see also Figure 1 below).

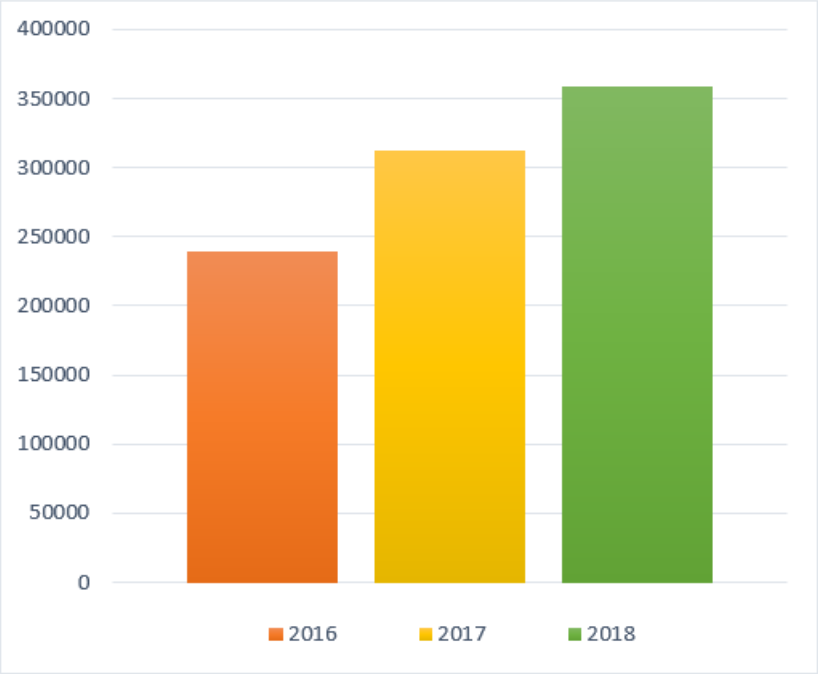


Figure 1. The total number of digital objects stored in EU LIS repositories

When looking at Figure 1, we can see that the European LIS repositories provided access to 360 000 various digital objects, while a number of accessed objects was growing year and year and the growth measure was 23 percent approximately. The greatest growth measure is observed between 2016 and 2017 and achieved up to 30.83 percent. However, we have observed a high level of dynamics as for the growth of several repositories. Many of them indicated more than 1000 percent growth between the observed years. The small systems are becoming the large ones and they are popular among visitors and number of visitors is growing continuously. However, that is also the main reason to carry on their safety.

JLIS.it 10, 2 (May 2019)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12545



## Section 2

In the next chapter, the testing was based on finding out if the HTTPS protocol is implemented in the login page or not. It was very simple. The systems that had the HTTPS protocol implemented were tested deeply using three independent online tools to achieve the highest rate of objectivity in the results. The repositories URLs were pasted sequentially into these online SSL Web server testers:

- by Qualis SSL Labs company,
- by Wormly company,
- by High-Tech Bridge company.

All these free online tools[1] perform a deep analysis of the configuration of any SSL web server. Every performed test is aimed to the depth of the analysis related to the current configuration of security certificates and supported cipher algorithms. It looks for vulnerabilities in the form of support of outdated technologies. Furthermore, the test simulates so-called handshake of various versions of operating systems, browsers and Java web technologies. As a result, the tests evaluate webs of repositories using the usual scale from A to F, which is also widely used in the academic environment. Partial steps, such as A- or B+ are also applied to achieve finer granularity of the results. A+ represents a better level of evaluation than A which is better than A- and so on (Formanek & Zaborsky 2017). Only Wormly company's tests use another metric for the results: a scale from 100% (the best) to 0% (the worst). In order to unify all the results, we have assigned the corresponding number of points on a scale from 10 to 0 according to the following table (Figure 2). Now, we are able to compare the results from all three tests.

| Results by | | The corresponding number of points |
|---|---|---|
| SSL Labs / HT bridge test | Wormly test | |
| A+ | 100% | 10 |
| A | 90% | 9 |
| A- | 80% | 8 |
| B+ | 70% | 7 |
| B | 60% | 6 |
| B- | 50% | 5 |
| C | 30% | 3 |
| D | 20% | 2 |
| E | 10% | 1 |
| F | 0% | 0 |

Figure 2. The table of assigned numbers of points

An example: if the repository gained score "A" from SSL labs, its actual test value was 9 points in this part of test. Next, it gained a better score "A+" from HT bridge in another test, so it got another 10 points. At last, the Wormly test scored only "49%", so it got only 4.9 points in the last part of the test. The average of these three tests is 7.97 points (9 + 10 + 4.9) / 3) for the tested repository in the given year. We compared only the average score obtained each year.

[1] URL addresses of the tests: https://www.ssllabs.com/ssltest; https://www.wormly.com/test_ssl; https://www.htbridge.com/ssl/. You can find here additional information as well as perform your own testing.

JLIS.it 10, 2 (May 2019)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12545



We tested all available functional repositories that OpenDoar listed based on our requirements (institutional repositories, from EU,… etc.). Tests were carried out in June 2016 and 2017. Now, we are able to add the new results from June of 2018 that could be affected by the GDPR regulation (which has been in the force from May of 2018). The results are very interesting, as it is shown later. Incidentally, we do not want to aim a negative attention to the specific systems, so we will present anonymous information only. We focus on the year-on-year difference between groups of the tested systems in this analysis. The objective of this analysis is also interesting because we can look at the summarized current results as well as the older data (see also Figure 3).

| Sequence number of repository | Average score calculated from the test results (SSL Labs, HT-bridge, Wormly tests together) | | |
|---|---|---|---|
| | 2016 | 2017 | 2018 |
| 1 | 9,90 | 9,90 | 9,33 |
| 2 | 9,30 | 9,67 | 7,97 |
| 3 | 6,20 | 9,53 | 9,67 |
| 4 | 9,33 | 9,33 | 9,47 |
| 5 | 9,33 | 9,33 | 9,10 |
| 6 | 5,97 | 9,10 | 9,57 |
| 7 | 8,97 | 8,97 | 9,33 |
| 8 | 0 (without HTTPS) | 8,63 | 9,20 |
| 9 | 6,67 | 7,97 | 9,47 |
| 10 | 7,63 | 7,63 | 4,93 |
| 11 | 0 (without HTTPS) | 7,30 | 9,33 |
| 12 | 5,53 | 6,87 | 7,30 |
| 13 | 4,00 | 6,67 | 7,73 |
| 14 | 2,63 | 6,63 | 7,97 |
| 15 | 9,53 | 6,20 | 9,20 |
| 16 | 4,97 | 5,97 | 7,40 |
| 17 | 4,20 | 5,87 | 9,00 |
| 18 | non-functional repository | 4,63 | 4,87 |
| 19 | 0 (without HTTPS) | | 7,30 |
| 20 | | | 5,60 |
| 21 | | | 6,87 |
| 22 | | | 8,30 |
| 23 | | | non-functional repository |
| 24-31 | 0 (without HTTPS) | | |
| 32-33 | non-functional repositories | | |
| 34 | The repository did not exist | The repository did not exist | 7,30 |
| Total average | 3,36 | 4,52 | 6,01 |

Figure 3. The test results

JLIS.it 10, 2 (May 2019)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12545



Every number shown in Figure 3 represents the average score of three tests performed step-by-step in 2016, 2017 and 2018. The numbers in one row correspond to the same repository. The tested repositories are listed anonymously and descending by gained score (higher is better).

## Results and discussion

The measurements provided in June 2016 indicated only 50 percent functional repository sites, which had been equipped with HTTPS protocol. That very low value indicated the alarm status because the LIS science deals with safety when providing data transfer and also communication. The actual systems run a risk closely related to the disaffection of access rights of users and administrators when not applying HTTPS protocol. As a result of that, those systems might be abused with fatal consequences. However, they might be damaged and the stored data could be as well as infected or deleted.

On the other hand, the safety characteristics related to the secured system did not indicate a high quality standard in 2016, either. The appropriate tests indicated that many systems got the worse evaluation marks, in spite of the implemented HTTPS protocol. However, also 50 percent of repositories without the above-mentioned protocol contributed to that very low average score 3.36 point. The test provided in June 2017 indicated a better situation, when more repositories equipped with HTTPS protocol were observed.

When having provided tests in June 2018, an increase of secured websites related to LIS repositories equipped with HTTPS protocol was observed, while 24 of 31 (74.2 percent) repositories contained that protocol. When comparing that with 2017, an increase more than 27.9 percent might be noticed, while an average point value is set at 8.01 (2.8 percent growth compared with 2017 and 15.4 percent compared with 2016). The total set of tested repositories (secured and not secured) indicates an expressive improvement between 2016 and 2018, it means from 3.36 to 6.01 point (78.9 percent).

## Conclusion and recommendations

Finally, we can say that many positive changes have been achieved related to EU LIS repositories environment within recent years. There has been a significant increase in the security of these systems. We anticipate that the improvements have been caused by the new European directive called GDPR, which entered into force on 25$^{th}$ of May 2018 in EU. However, more than a quarter (25.8%) of systems remains also unsecured. Based on GDPR (inter alia) all of the companies should ensure that the transfer of sensitive data might not be misused in the computer networks. These data contain also login credentials and other sensitive data used by various types of systems such as digital repositories. Reliable and powerful end-to-end encryption is an appropriate solution for information transfer.

At the end, it is very important to regularly monitor and test the level of web security (not only) of academic and institutional repositories. It is necessary to respond quickly to the feedback reports, which are usually part of the evaluation reports – the outputs from the three tests mentioned above. Just follow the instructions in the evaluation report: There is usually a list of reasons why a particular site was awarded a lower rating. The administrator can immediately make an appropriate action: modify the configuration of the web server, specify a set of supported cipher's sets. This is very easy to do. It can be done by an advanced user, too. You can achieve significantly higher/better ratings by

JLIS.it 10, 2 (May 2019)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12545

simply restricting weak ciphers and protocols by changing your web server settings. On the other hand, when you strictly support only the most up-to-date set of encryption ciphers, your repository may not be available from older OSs with older versions of Internet browsers. So, changes need to be tested and implemented with caution.
For example, we recommend using the stronger ciphers[2] in a configuration file (/etc/tomcat8/server.xml) of Tomcat web server running on Linux-based operating systems. A key step is to delete the weak ciphers from this file and to support only modern and stronger algorithms.[3]

## References

Adnan, A. *et al.* 2017. "Compromised User Credentials Detection Using Temporal Features: A Prudent Based Approach". Proceedings of the 9th International Conference on Computer and Automation Engineering. 104–110. ISBN: 978-1-4503-4809-6. Doi: 10.1145/3057039.3057051.

Duraspace. 2018. DSpace manual [online]. 2018-06-27. Accessed November 09, 2018. Available online: https://github.com/DSpace/DSpace/releases/download/dspace-6.3/DSpace-Manual.pdf.

Enterpreneur Europe. n.d. "HTTP vs. HTTPS: What's the Difference and Why Should You Care?" [online]. Accessed September 06, 2018. Available online: https://www.entrepreneur.com/article/281633.

Formanek, M. and Zaborsky, M. 2017. "Web Interface Security Vulnerabilities of European Academic Repositories" [online]. Accessed November 15, 2018. Available online: https://www.liberquarterly.eu/articles/10.18352/lq.10178/.

Kalman, G. 2018. "10 Most Common Web Security Vulnerabilities" [online]. Accessed September 15, 2018. Available online: https://www.toptal.com/security/10-most-common-web-security-vulnerabilities.

Sugavanesh, B. *et al.* 2013. "SHS-HTTPS enforcer: enforcing HTTPS and preventing MITM attacks." November 2013 ACM SIGSOFT Software Engineering Notes: Volume 38 Issue 6, November 2013, 1–4. Doi: 10.1145/2532780.2532802.

[2] SSLEnabled="true" sslEnabledProtocols="TLSv1,TLSv1.1,TLSv1.2"ciphers="TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256, TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA,TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA384, TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA,TLS_ECDHE_RSA_WITH_RC4_128_SHA, TLS_RSA_WITH_AES_128_CBC_SHA256,TLS_RSA_WITH_AES_128_CBC_SHA,TLS_RSA_WITH_AES_256_CBC_SHA256, TLS_RSA_WITH_AES_256_CBC_SHA,SSL_RSA_WITH_RC4_128_SHA".

[3] There are more information about this topic on specialized security web sites: https://ssl.comodo.com/support/ssl-technical-faqs/how-to-disable-weak-ciphers-in-tomcat-7-8.php or https://www.sslshopper.com/article-how-to-disable-weak-ciphers-and-ssl-2-in-tomcat.html.

JLIS.it 10, 2 (May 2019)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12545

Tennant, R. 2017. "Current Cities" [online]. 2017. Accessed September 21, 2018. Available online: http://currentcites.org/2017/cc17.28.2.html.

The Https-Only Standard. n.d. [online]. Accessed September 20, 2018. Available online: https://https.cio.gov/everything/.