Zranitelnosti institucionálních repozitářů

Kovářová, Pavla
2011

Seminar on Providing Access to Grey Literature 2011: 4th year of the seminar focused on storage and making the grey literature accessible, 25. 10. 2011 [online]. Praha: National Technical Library, 2011. Available at WWW: http://nusl.techlib.cz/Sborniky. ISSN 1803-6015.

VULNERABILITIES OF INSTITUTIONAL REPOSITORIES

PAVLA KOVÁŘOVÁ

kovarovap@gmail.com

KISK FF MU

Abstract
The aim of this paper is to introduce the problem concerning the threat to the data for the purposes of development and maintenance of institutional digital repositories. The paper points out to potential sources and targets of threat and security options. Therefore the goal is to raise awareness of stakeholders regarding the threatening situation. This is achieved by illustrative examples of threats that can be encountered in real situations and in particular by references to more sources of information concerning the topic.

Keywords
Risk Analysis – Digital Repository – IT policy – Security

1. DEFINITION OF TERMS AND TOPIC

The importance of documents is well known from the history, however such documents can only be used if collected, organised, managed and accessed in a proper way. This has been long and much pressing matter for people. The development of information and communication technologies has caused that the access to the information recorded in documents is more and more often seen as self-obvious, sometimes almost as a fundamental human right[1]. Naturally, this is a simplification, and the one that can easily be questioned, but it is not the aim of this paper. Therefore it would suffice to say that more and more pressure is being put by the public to make documents that were created, at least partially using public funds, also accessible to the public. Moreover, to make the work more effective, in particular in large companies, it is necessary to make accessible the required documents regardless the geographical distance from the location where the documents are stored. All of this is reflected in the increasing number of institutional repositories.

Before we can address the very topic of this paper, it would be convenient to outline its scope and define the basic term for exact interpretation of its content.

[1] FRANK, La Rue. United Nations : Human Rights [online]. 16 May 2011 [cit. 2011-09-15]. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue. available z WWW:

<http://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf>.

As it is clear from the title, this paper does not deal exclusively with grey literature, as much as it is a main topic of the workshop, where it has been placed. The topic has been extended for a simple reason. The literature itself is problematic, especially in the copyright area, which can be evidenced and sufficiently described by papers focused on the grey literature in the legislation context. The major problems does not have to necessarily come into the light when the law is breached by the administrator or the user of documents, but also when the documents are stored, processed and searched in both traditional and digital repositories. Thus the attention will be paid mainly to digital institutional repositories, particularly the ones in the field of science and research. Then integral part of this system is both grey and white literature.

This paper is mostly concerned with the introduction of the issue of threats to the data for the purposes of development and maintenance of institutional digital repositories. Due to the scope of the topic, the core of this paper comprises of several examples of vulnerabilities of institutional repositories, which, needless to say, are identified by both their developers and administrators. The aim is not only to show the threats but their solutions as well, even if the 100% security cannot be achieved. By references to other resources this paper shall point out to measures for the risk reduction. The purpose is not to discuss the topic in more detail, as the application of the solution will largely depend on the particular system.

Few terms, that may seem too strong, such as vulnerability or threat, are used in the text. However, this is a professional terminology. It was clearly defined already in the book Informační bezpečnost (information security)[2] by Požár that I can strongly recommend to everyone who is interested in the publications summarizing the fundamentals of IT system security. And now to the very definition of the terminology used:

"The Threat is a circumstance, occurrence, force or persons, the operation (activity) of which may cause a damage, loss, loss of confidence or value of an asset. (…)

The Risk is a probability, with which the value of an asset will be destroyed or damaged by the operation of a particular threat working against the weakness of this value. (…)

The Attack, which we also call a security incident, means either an intentional use of the weak spot, that making use of the weak spot to cause damage/loss of IS assets, or unintentional action result in the damage to the assets. (…)

The Vulnerability is a deficiency or a weak spot of the security system that can be used by the threat resulting in the damage or loss of the assets."[3]

Whether a new digital repository is built or a proven solution is used, it can never be said that the system would be free from vulnerabilities. Every system security has sooner or later a weak spot, even though in some cases only after several centuries, as in the Vigenèr's

[2] POŽÁR, J. Informační bezpečnost. Plzeň : Aleš Čeněk, s.r.o., 2005, s. 309.

[3] Ibid, page 37-38.

"unbreakable" code. Owing to the fact that ICT development and involvement of more and more people who seek vulnerabilities we can more likely speak of hours, days or month at maximum, during which the vulnerability is discovered.

In case of digital repositories the vulnerabilities arise mostly as errors during the analysis, development or implementations of solutions and can have form of both intentionally and unintentionally created methods of use that were not intended by the sponsor. However, because the repository is not only composed of a technical solution, but of documents and users as well, the vulnerabilities can appear also there. Therefore it is primarily necessary to keep in mind the human element, which is seen as the weakest link of the security.

Despite the fact that the vulnerabilities cannot be avoided, we cannot give up the benefits resulting from ICT, as then we will have to resign to the development of the institution, which will render it insignificant. The security of digital information system has been dealt with from the very beginning. The vulnerabilities are rapidly solved with institutions that cannot do without such systems, which are particularly large multinational corporations. These also have enough funds to invest and seek solutions, which are later made public in a certain level of detail. Thanks to this, the solution can also be used by institution that cannot afford such investments. Therefore if a state-of-the-art security is maintained, the institution will be protected against most of the attacks.

For the analysis of the information system security, risk assessment is used, which is again defined by Požár as "a process of assessing the threats working against *the information system with an aim to define the level of risk, to which the system is exposed. The purpose is to establish whether the security measures in place are sufficient to reduce the probability of the occurrence of the damage to an acceptable level.*"[4] Risk management therefore takes into account the probability of an attack and the effectiveness of the protection as often the vulnerability is identified, but the costs to remove the vulnerability would be higher than the value of the asset at risk. Therefore when decision to implement countermeasures is considered, the so called ALE (Annual Loss Expectancy), that is the expected annual lost, is calculated. The details can be found in publications on management, risk management, or information systems security, as the case may be.

Risk management is also an essential but not a single element of the information audit - not entirely unknown to those who follow the activities of the NUŠL project. The NUŠL project carried out the information audit in 2009 and 2010 and the experience and results [5] are published. NUŠL is specific, but some vulnerabilities are common to other repositories as

[4] Ibid, page. 37.

[5] KARLACH, Petr. Repozitáře šedé literatury. 1. vydání. Zlín : VeRBum, 2010. The NUŠL audit using the DRAMBORA tool, page 123-132. Available at WWW: <http://nusl.techlib.cz/images/Book.pdf>. ISBN 978-80-904273-5-8.

National repository of grey literature [online]. 2010 [cit. 2011-09-15]. Audit of the National repository of grey literature (NUŠL) in NTK using the DRAMBORA tool: Second audit in 2010. Available at WWW: <http://nusl.techlib.cz/images/AUDIT_2010.pdf>.

well. Other developers and administrators of digital repositories can make use of the experience and advice of the NUŠL project team for their own audits.

2. CLASSIFICATION OF POSSIBLE APPROACHES CONCERNING THE TOPIC

The last general part of this paper constitutes a bridge from the theoretical base to the promised examples of vulnerabilities that will best describe the whole topic and will help us to picture the topic in practice. It mostly deals with the factors that play a role when a vulnerability of digital institutional repository arise or is abused. These are transparent in the classification shown using the mental map in Fig. 1 (Vulnerabilities of Institutional Repositories).

Each weak spot and threat can be described from different aspects. The form and way of how the source is manifested in the attack are of crucial importance. With the help of classification we can also take a look at the repository as a whole and identify the majority of vulnerabilities. Its use thus involves risk assessment.

To be more specific of what the mental map shows the individual parts with brief description and their forms in real life and their impact on the security will be presented. At the same time, it will be indicated that the vulnerabilities of digital institutional repositories do not somehow aggravate the weak spots of institution, because if this system was not used these would manifest in different forms.

When identifying the vulnerability, all types of threats must be taken into account. The natural threats such as floods have occurred repeatedly in recent years. The solution might be namely the appropriate back-up strategy more increasingly connected with the so called cloud services. The physical threats are mostly connected with technical devices that can, but does not necessarily need to; put at risk data, programs and services. Another example is a theft of hardware. If an emphasis is placed on data, data back-ups can be used again, and physical measures such as locks, security guards and so on must be used to secure the device. The natural and physical threats can also affect the traditional repositories where the security and data back-ups in particular, are more complex. The technical and physical threats have one thing in common, in both cases the data and devices themselves can be at risk. However in vast majority in cases these are classic defects. The security is the same as with the physical threats. The last but most problematic type of threats seems to be the human threats that can either be intentional or unintentional. Again, these can be oriented against data or hardware, but also against administrators and users. The hardware security is no different from the previous cases; however the data and human targets pose more problems which relates to the human inventiveness. Here we cannot simply mention one method of protection, as all items given in the area of security are of key importance, and can take various forms.

If we stay with human threats, their source has a crucial role. It is also important for other three types of threat, but from the security aspect, it brings us down to the very level of specific measures, e.g. when the back-up strategy is chosen. However, as it has been said above, the human threats can take various forms and therefore the source can help to identify

the system security. The crucial point is whether the threat comes from outside or inside. While the threats from the outside are practically always intentional aiming to gain benefit or harm, the internal threats can be both intentional and unintentional. As the internal threats come from those who must work with the repository, it is harder to be protected, because it is necessary to find the thin line that would still enable users to use the repository, however without abusing it. The research shows that the threat from the inside comes more often than from the outside[6]. The solution can be seen in the right setting of software and hardware, with which the data management is closely related. In the external environment practically anybody can be a source of threat. Most of time completion springs to mind but it can also be a former employee, an enemy to someone who is responsible for given data or repository, or even a person who just wants to test his/her abilities to attack. For protection, the same means are used as in the event of an internal source, accompanied with back-ups, physical security and specialized security applications. No matter what the source of threat, education of repository users and their responsible conduct can have major impact.

The form of an attack from the security aspect is rather a specific one, and a more detailed description should be part of the introduction. Still it might be appropriate to include it in the mental map, as it allows us to show the variety of attacks. Concerning the digital repositories, suspension means that the repositories are not functional, and this can easily be caused by natural, physical or technical threat, but it can also be an attack through the Internet. A typical example are the so called DoS (Denial of Service) attacks when the repository is intentionally overloaded with so many requests that the repository cannot respond and the access is suspended even for authorized users. These attacks are most often targeted at programs and services, or hardware. We do not need to go into much detail as far as the wiretapping is concerned, although it should be said that attacks to illegally obtain information occur more often than the attacks to harm the data, the amount of which decreases, which is mostly given by financial motivation with the alteration and adding the information to the data being the two types of potential damage to the data.

The conflict with the law is mentioned because then the minimum necessary security of vulnerabilities can be established, as for almost everyone who builds an institutional repository, the non-applicability of law is crucial. This and other papers in the NUŠL workshops have in much detail described the key spots regarding the Copyright Act[7].

[6] In the Czech environment, the following articles mentioned it:

ITC Security [online]. July 8, 2011 [quot. 2011-09-15]. Za krádeží a únikem citlivých informací stojí nejčastěji zaměstnanci (Most of the time, the employees are responsible for information thefts and leaks). Available at WWW: <http://ictsecurity.cz/security-bezpenost/za-kradei-a-unikem-citlivych-informaci-stoji-nejastji-zamstnanci.html>.

ITC Security [online]. February 3, 2011 [quot. 2011-09-15]. GFI Software: V roce 2011 vzrostou bezpečnostní hrozby prostřednictvím výměnných paměťových médií (In 2011, security threats will increase through exchangeable storage media) Available at WWW: <http://ictsecurity.cz/security-bezpecnost/gfi-software-v-roce-2011-vzrostou-bezpecnostni-hrozby-prostrednictvim-vymennych-pametovych-medii.html>.

Similar results are given in POŽÁR, J. Informační bezpečnost. Plzeň : Aleš Čeněk, s.r.o., 2005, s. 309. str. 60.

[7] The Czech Republic. The Act on copyright and related rights acts and on change of some other acts (Copyright Act), In Coll., the Czech Republic. 2000, 36, p. 1658-1685.

Information, the protection of which is regulated by several laws, in particular the personal data[8] that can be part of the user profiles or documents in repositories, and classified [9] or secret information, which can have a form of trade secrets[10], were mentioned. In a way, the documents violating copyrights or the protection of the above mentioned information, represent illegal content. However, such content can have other forms, e.g. content promoting racism[11], depicting illegal forms of pornography[12] and so on. Or, as the case may be, it can be an inappropriate content, which is illegal and is made accessible to certain group of people, e.g. pornography and children[13]. The applicability of law then depends on the characteristics of the repository. In detail, this is dealt with by the book Právo na internetu (LAW on the Internet) [14]. The problem with inappropriate and illegal content concerns mostly repositories where documents can be stored by other users than system administrators or by those who study such documents as subject of science.

The last two parts of the map that is the target of the attack and security were reflected when describing the previous aspects, therefore they are not necessary to be dealt with anymore.

The security against intrusion in the repository through identified vulnerabilities is easier than the blind application of the security measures. As a tool for seeking possible protection and improvement the level of security that can play role for the liability by law (e.g. ensuring the security of personal data), several documents were prepared. The following as of special importance:

- ISO standards:
    - ČSN ISO/IEC 27000 – Information Security Management System –Overview and Vocabulary,
    - ČSN ISO/IEC 27001 - Information security management systems — Requirements,
    - ČSN ISO/IEC 27002 - Code of practice for information security management,
    - ČSN ISO/IEC 15408 – Criteria for IT security assessment,

[8] The . The Act No. 101/2000 Coll. on personal data protection and on change of some other acts. In Coll., the Czech Republic. 2000, 32, p. 1521.

[9] The Czech Republic. The Act No. 412/2005 Coll. on protection of classified information and security capacity. In Coll,. , the Czech Republic. 2005, 143, p. 7526.

[10] The Czech Republic. The Act No. 513/1991 Coll. Commercial Code. Coll., the Czech Republic. 1991, 98, p. 2474.

[11] The Czech Republic. The Act No. 40/2009 Sb. Criminal Code. Coll, the Czech Republic. 2009, 11, p. 354. Section 355, Section 356, Section 403 and 404.

[12] Ibid, Section 191 and 192.

[13] Ibid, Section 191.

[14] POLČÁK, Radim. Právo na internetu : spam a odpovědnost (Law on the Internet: spam and responsibility) ISP. 1. vydání. Brno : Computer Press, 2007. 160 s. ISBN 978-80-251-1777-4.

- ČSN ISO/IEC TR 13335-1 – Guideline for IT security management, Concept and models for IT security management,
- ČSN ISO/IEC TR 13335-2 - Guideline for IT security management, IT security management and planning,
- ČSN ISO/IEC TR 13335-3 - Guideline for IT security management, Techniques for IT security management,
- ČSN ISO/IEC TR 13335-4 - Guideline for IT security management, Selection of protective measures,
- ISO/IEC TR 13335-5 - Guidelines for the management of IT Security, Management guidance on network security

- Documents for the assessment of the system credibility:
  - TCSEC - Trusted Computer System Evaluation Criteria, the so called Orange Book: an old document but serves as a basis for a newer one
  - ITSEC - Information Technology Security Evaluation Criteria: followed by ITSEM evaluation manual
  - CTCPEC - Canadian Trusted Computer Product Evaluation Criteria

From the practical point of view, the methodologies and software solutions, such as CRAMM and Cobra are more beneficial. In the field of repositories, the DRAMBORA tool, which was used for the above mentioned NUŠL audit, must be mentioned. However, as opposed to NUŠL, the institution should limit the publication of results, because by doing so it gives out its weak spots and thus points out to vulnerabilities.

By saying so, we can sum up the part that introduced the vulnerabilities from several viewpoints and sources, the aim of which is to help the developers of information systems and people who are responsible for them to ensure that the system is secure enough for the needs and options.

3. EXAMPLES OF THREATS, VULNERABILITIES AND PROTECTIVE MEASURES

In order to describe various vulnerabilities and their security in more detail, the examples of the three most frequent types of security issues concerning the repositories are given. It is the violation of secrecy, integrity and availability.

3.1 VIOLATION OF SECRECY

When secrecy is violated the information is disclosed to someone who should not have access to it. A typical example is a theft of data for competition purposes. For more detailed description, it is appropriate to use the aspect given in the above mentioned mental map.

It is a human threat, which can be intentional if led directly or unintentional if led through the unaware user of the repository. At the same time, the so called social engineering is used, which increasingly accompanies various attacks. This means that attacker persuades the victim to do something what will bring the attacker the benefit and then the attacker harms the victim. By using social engineering the secrecy may be compromised even without the knowledge of the abused user or any other persons. Apart from that the attack can be led using coercion, in particular in competition environment in a form of corruption, blackmailing and so on. When led directly or through intermediaries, the attack can occur also in physical environment, for instance through a break-in, or using ICT or the Internet where it is not difficult and find the co called spyware.

It follows from the above that the primary source of target is competition, but the last link, which abuses the vulnerability, may be a competitor, attacker or user or administrator of the repository. Both forms, being the wiretapping, and target, being the data of the attack, are also clear. From the conflict of law perspective, the proprietary information that can be part of the stolen data, may come to mind.

- When securing the system, the data administration in a form of a secured access is of key importance. In particular, this means authentication that is verification whether the user is the one who he declares to be and authorisation, which means determination of the right to access to specific data or services. At the same time, it may be appropriate for larger repositories to limit the access also to the authorized users so that they could access only data on the need-to-know basis. Security applications protecting against all kinds of mallware and firewall securing the data streams, or more advanced instruments, which may however be problematic due to their high price, are also a must. The theft of data might also occur in the physical environment; therefore the countermeasures include the physical security. As with other threat, for this threat the education and responsible conduct off the users is crucial as only such conduct can protect the users against some social engineering techniques.

The problem of data theft is not limited to digital repositories, even if the risks are higher with them. The same threat may be caused by the data printed or recorded on media (CD, flash disk and so on). Therefore these possibilities must also be taken into account and it is necessary to inform the user of vulnerabilities of both traditional and digital repositories and of procedures of how to respond to them. This should be part of the so called information policy, which should be in place in every institution.

-

3.2 VIOLATION OF INTEGRITY

- When integrity is violated, value is altered or added to the database, which defines both the form of an attack and its target that is again the data. The problem is that the user will access improper data and while using them the user suffers harm. As a result, the administrator of the repository is harmed as he loses credibility. The administrator may be at risk either directly if integrity is violated in way that is provided for by the law. Then basically any conflict with law, especially concerning copyrights, proprietary information, or inappropriate and illegal content, may occur.

- When integrity is violated, it is again a human threat that can be intentional if it aims to harm the user of the repository administrator or unintentional if there is only a failure without the intention to harm. The unintentional threat comes from the inside, the intentional from anywhere.

- As far as security is concerned, back-ups might be a proper solution that can help to recover the proper data, unless integrity violation has been identified. The physical security can prevent the attacks from physical environment; in particular it is appropriate to make the data accessible only from a certain point. Other security methods given above are of key importance to prevent the violation of integrity, retrospectively there is not much they can solve.

-

3.3 VIOLATION OF AVAILABILITY

- The last type that should be dealt with in more detail is a violation of availability, the classic example of which are the DoS attacks described above. They mostly take form of suspension while it targets the programs and services. It is almost entirely the intentional human threat with its source being the competition of attacker and therefore it comes from the outside. The protection is not an easy one and proper setting of hardware and software and some security applications might help. However, there are often cases when the target of such attack is helpless and can only wait until the attack stops.

-

4. SUMMARY

The basic vulnerabilities of digital institutional repositories were introduced. Using the examples from real life the topic was described in more detail and put into context of the day-to-day life of administrators and users of institutional repositories. This topic is seldom dealt with in the professional literature, which mostly describes the vulnerabilities more generally focusing on the information systems of all kinds. Naturally, some attention should also be paid to a specific field of institutional repositories, namely due to their growing importance and number.

The paper outlined the risk that may occur in connection with institutional repositories and pointed out to their possible limitations. Some indication was also made to the fact that in spite of the vulnerabilities that institutional repositories may have, even though these cannot be entirely eliminated, the traditional repositories have the same or other weak spots. Therefore not introducing the digital repositories is not a solution to support and protect the institution, but on the contrary, making use of them while respecting the potential threats and taking such countermeasures that might be appropriate is an answer.