CZĘŚĆ I
Cyberbezpieczeństwo, ataki i przeciwdziałanie cyberprzestępczości

Mirosław Kutyłowski

Rozdział 1.
Contemporary Damocles sword: Cryptographic tools facing anamorphic schemes

1. The growing role of trusted devices

The growing role of IT systems in all areas of social activity creates tough challenges in both the technical and legal sense. One of such challenges is the role of natural people who interact with the systems. For decades, we have been using the concept of a **user**: as an individual who interacts with the technical system, initiates some processes, and receives digital feedback from the system. This point of view describes very well the reality decades ago, when all we had were island IT systems, with very sparse interaction between each other and with their own sets of users.

The situation has changed significantly during the last two decades; not only has a significant number of services and processes been moved to the digital ecosystem, but also the number of users has grown to include all people, except for those **digitally excluded**. Therefore, we should now refrain from using the term "user" and, instead, talk about a **digital person** who is related to a physical person.

Although a digital person can perform activities such as participating in advanced authentication protocols or signing digital documents with the use of strong cryptographic schemes, the role of the physical person behind a digital person is merely to perform some marginal actions, such as clicking a "yes" button or providing a static PIN to a system which performs crucial operations. In order to ensure a minimal level of control over a digital person by its physical counterpart, the processes initiated by the digital person are quite frequently run within a separate physical unit, such as a smart card. In

this way, the physical person has the opportunity to switch off the device, simply by inactivating it through withdrawal from a card reader, encapsulating it in a Faraday cage, etc. On the other hand, we can pay more attention to the design and verification of the security features of isolated devices of this kind.

Currently, many IT systems use log-in and password authentication. As a result, a user is required to remember a growing number of passwords, which turns out to be impossible if the passwords have to be strong and independent of each other. There are a few solutions that reduce the burden on the user, but all of them also take away control from him. Again, these solutions shift the core operations from the physical person to their digital agent. For example, password managers installed on smart devices allow a user to keep all passwords in an allegedly secure vault, the access to which is controlled by a single password. On the upside, this approach prevents the creation of weak passwords and/or reusing passwords, but the control is taken away from the user – in fact, a breach into a password manager enables impersonation of the user in a large number of systems. Although the overall security of the operating systems installed on the devices which hold these password managers is limited, this solution may only be regarded as a risk-limiting procedure.

A different concept has been developed for German personal identity cards (cf[1]). The German electronic identity card, i.e., the so-called nPA, has been equipped with a functionality called Restricted Identification. The idea was that the nPA generates a unique password for each so-called *domain*, i.e., a target IT system. The generation process is based on a cryptographic function implemented on a secure device. The advanced cryptographic properties of this function prevent others from guessing the password and make it impossible to link two passwords in different domains. The last property means that it is infeasible to tell whether two passwords have been generated by the same device. On the upside, the authentication process is delegated to the nPA which is a trusted device with a native operating system that performs only a well-specified and closed list of functionalities. On the downside, this is a black box device.

The idea behind Restricted Identification was not the only concept of this kind. BSI, a German federal authority, developed the scheme of Pseudonymous Signatures[2], where the user can create not only an unlinkable pseudonym for a domain, but also digital signatures corresponding to that pseudonym. Again, while on one side this contributes a lot to privacy-by-design and to protection of personal data by technical means, the responsibility is shifted to a secure black box device. Unfortunately, the overall concept is based on the

[1] BSI TR-03110 eIDAS Token Specification, https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Technische-Richtlinien/TR-nach-Thema-sortiert/tr03110/BSITR03110-eIDAS_Token_Specification.html (last accessed: 1.4.2023).
[2] Ibid.

assumption of a trustworthy issuer of personal identity cards . There are multiple ways in which such an authority can misbehave – among others, by leaking information that enables linking of pseudonyms.

This trend of shifting the basic functionalities of citizen identification and authentication to trusted devices has also been indirectly reflected by the eIDAS regulation[3]. Assurance levels of the electronic identification schemes from Article 8 of this regulation refer to 'electronic identification means'. In the current state of security technology, it is hard to imagine any solution that would implement 'electronic identification means' and that would not be based on secure devices that implement advanced cryptographic functionalities. Otherwise, it would be extremely hard to fulfill the requirements even for the assurance level "substantial".

2. Anamorphic schemes

The primary purpose of black box devices is to protect the internal memory of a device, including all ephemeral values created by the device during the execution of the protocol. Although the ultimate goal is to protect the private keys implemented on the device, hiding other values is almost always necessary, as they may indirectly leak the values of the secret keys. This is not surprising, since designers of cryptographic schemes silently assume that the only information available to an observer is the official output of the scheme. The exceptions are 'leakage resistant schemes'; however, even then we are talking about a bounded leakage. Moreover, it turns out that it is necessary to prevent information leakage by examining execution characteristics, such as execution time and energy usage. A whole branch of cryptoanalysis is based on such side-channel information.

This leads to a situation where the only information available to an observer of a black box device is the official output of a cryptographic scheme. If there are deviations from this ideal situation, then it is considered a weakness of the device. Unfortunately, the security means discussed above turn out to be a double-edged sword, as they make room for anamorphic schemes applied for evil purposes. The idea behind an anamorphic scheme D', designed on top of a scheme D, is that, for an external observer, D' behaves exactly like D and one cannot distinguish between the black box devices that implement D and D'. On the other hand, D' may implement additional hidden functionalities.

The goal of an anamorphic encryption scheme is to create anamorphic ciphertexts. If the receiving party has no appropriate decryption key, the anamorphic ciphertexts look like regular ciphertexts created with the basic scheme. Even if the encryption/decryption keys of the basic scheme are surrendered to

[3] Regulation (Eu) No 910/2014 Of The European Parliament, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32014R0910&from=PL#d1e1233-73-1 (last accessed: 1.4.2023).

an inspector, he cannot see any difference and suspect that the output comes from a device implementing anamorphic encryption. On the other hand, if the recipient of the ciphertext knows the appropriate decryption key, called a *dual key*, then he may extract the hidden plaintext. There are no limitations for hidden plaintexts, apart from their size.

Anamorphic encryption was presented in 2022 [4] as a remedy against dictator regimes that may require (by national laws) to surrender the keys used for encryption and decryption. The excuse may be, for example, the fight against terrorism. In this way, an anamorphic encryption scheme could protect dissidents from dictators: If a request is made to surrender the keys, there will be nothing that could endanger the dissident: the hidden ciphertext will not only remain confidential but its entire existence will be hidden.

The downside of anamorphic encryption schemes is that the capacity of the additional channel is limited: this results from the basic principles of information theory. Nevertheless, it turns out that many standard encryption schemes could be reshaped in this way so that there is sufficient capacity to transmit the standard secret keys.

A dual key can be easily established between the party that created the ciphertext and the receiver. Even if the ciphertext is created with the public key of a party X, it is possible to establish a dual key with another party Y, without even having any contact with party Y.

The major difference between these recent results and the results on subliminal channels is that anamorphic schemes should be based on widely used standard schemes. This is necessary, since the use of any exotic scheme enabling a subliminal channel would be enough for the dictator to begin persecuting the person using such a scheme. Anamorphic schemes need not be limited to encryption[5]. Similar constructs can be based on cryptographic primitives such as:

1) digital signature;
2) online authentication;
3) establishing a session key.

In particular, one can easily embed anamorphic ciphertexts in transactions of cryptocurrencies. Due to the distributed ledger principle, such transactions would be made available to any participating party. Thus, even if the dictator knew that party A was inserting hidden ciphertexts into transactions, it would be impossible to detect who is the recipient of the relevant message.

Anamorphic schemes are enabled by the intensive use of (pseudo)random number generators by cryptographic schemes. Randomness is the simplest

[4] Cf. *G. Persiano*, *D. Hieu Phan, M. Yung*, Anamorphic Encryption, pp. 34–63.
[5] Cf. *M. Kutyłowski*, *G. Persiano*, *D. Hieu Phan, M. Yung*, *M. Zawada*, The Self-Anti-Censorship Nature of Encryption.

way to prevent the prediction of the internal state of a secure device. There are only a few examples where computations are deterministic. A prominent example is EdDSA, where the ephemeral value for signature creation is derived as a hash value over the message to be signed and an extra secret key. However, even then, it is easy to implement an anamorphic channel, as the black box device can replace the procedure for generating the ephemeral value. In this case, cryptography would help to hide the hostile change. Even RSA encryption, deterministic in its textbook form, enables an anamorphic construction, due to the padding and symmetric key (in case of the hybrid ciphertext).

Removing the playground for anamorphic constructions is impossible for two reasons: first, it would be impossible to replace the already deployed algorithms on such a wide scale. In many cases, there are no technical conditions for performing such a rapid roll-out, if the schemes are embedded in hardware. Second, alternative designs are still in their infancy. So far, there has been no R&D focus in this direction, comparable to the development of post-quantum schemes. Moreover, many researchers claim[6] that efforts to stop anamorphic schemes and other similar solutions are doomed to fail.

3. Attack vectors

Although anamorphic schemes may serve good purposes, as indicated in the literature[7], one can show that anamorphic schemes are ideal for implementing treacherous black box devices. The primary examples are malicious authorities or enterprises providing allegedly secure black box devices to implement core security functions. These devices could pretend to implement well-designed and widely accepted standards, whereas, in reality, anamorphic schemes may be deployed there.

The goal of the adversary might be to obtain data from the malicious device, which would normally never be exposed. Let us list some major threats:

1) **Leaking the symmetric key chosen for hybrid encryption:** The addressee of the leakage could be a third party, not necessarily the provider of the devices. An example of such a third party would be the political police of the dictator. Leaking the hybrid symmetric key selectively dismounts the standard cryptographic protection – allegedly confidential communication is actually available to the designated eavesdropper. The scope of this threat is substantial, as hybrid encryption is fundamental for secure communication on the Internet.
2) **De-anonymization in anonymous communication:** in the case of privacy-protecting communication schemes (such as TOR), the treacherous device may implant hidden fingerprints into pseudonymous messages,

[6] Cf. *G. Persiano*, *D. Hieu Phan, M. Yung*, Anamorphic Encryption, pp. 34–63.
[7] Ibid.

so that privacy protection becomes ineffective against the adversary. The leakage may affect both the source and the destination of the message.

3) **Signature forgery:** a hidden channel may be implemented in a signature, which will result in the addressee of this hidden information getting the private signing key. This kind of attack is devoted to the case where the signing key is not handed to the user together with the device, but is later securely created on request of the user. There are secure schemes that enable this; however, the protection becomes an illusion, if such a carefully created key is leaked to the adversary.
4) **Impersonation:** an anamorphic scheme can leak credentials and ephemeral values that might enable the attacker to impersonate a user in a separate session or to hijack the session initiated by the user. In the latter case, the leakage might be selective: secret authentication keys need not be revealed to the designated attacker[8], e.g., as in the case of password authentication PACE on European personal identity documents.
5) **PIN leakage:** an anamorphic scheme may be used to disclose the PIN number to the device. This would open doors for attacks based on the use of the device by the adversary without the user being aware of it. All types of "lunch break attacks" could be launched (e.g., a device is left in the office during lunch break; the insider attacker activates it with the legitimate PIN and uses it for rogue purposes).

A significant property of all these attacks is that they could be latent. If they are passive, they could be completely undetectable for the user (e.g., de-anonymization of allegedly anonymous communication or breaching confidentiality of a secure channel).

A great problem is the false sense of security: A user holding a device that allegedly implements an *approved* cryptographic scheme (e.g., within the FIPS framework) may become less careful in their actions. This is exactly what the dictator hopes for.

4. The right to security and verification

Gradually, it is becoming evident that personal security in the cybersphere should be recognized as one of the fundamental human rights. To some degree, this approach is already represented in the Charter of Fundamental Rights of the European Union and in GDPR. Personal data protection is not considered to be an abstract concept; it appears repeatedly in the context of fundamental rights and freedoms of the data subject. The real reason for personal data protection is definitely not the abstract concept of privacy – po-

[8] See *M. Kutyłowski*, *A. Lauks-Dutka*, *M. Zawada*, ICAO Travel Documents in Hands of a Dictator, http://kutylowski.im.pwr.wroc.pl/articles/CECC22-talk.pdf (last accessed: 1.4.2023).

tentially disrupting social contacts. The real reason is the threat of misusing personal information.

Unfortunately, the right to security, in particular in regard to security in cyberspace, has not been included in the Charter of Fundamental Rights of the European Union. Similarly, the recently formulated *right-not-to-be-deceived*[9], has not been adopted (yet) and is not a fundamental concept. This leads, for example, to freedom to deceit individuals in many areas, such as business and political campaigns based on deceptive data. There are examples in the legal systems outside the European Union where the misinforming party becomes effectively responsible for the deception. A good example is Taiwan where, when promoting medical products, the advertising party is obliged to keep its promises made to consumers. The practice in Poland is exactly the opposite: the media are full of false (and even dangerous) advertisements of a similar kind.

Whereas one could claim that the right to security in cyberspace is one of the fundamental human rights, this should concern, in particular, security of the devices implementing the fundamental security functions. As in the case of personal data protection, the relevant properties of these devices should be demonstrable. This stems in particular from Article 5 of GDPR, as devices typically process data related to an identifiable person. Even if a physical person uses the device to create a pseudonym and operate under the said pseudonym, appropriate protection of the rights and freedoms of such a person should be guaranteed. In particular, this means protection against de-anonymization via anamorphic setups.

Having agreed that security of a black box device must be demonstrable and not only declared, the question arises of how to convert the declarative right into an effective technical solution. GDPR is only the first step, providing a firm response to all attempts to neglect the issue or to challenge the right to verification – on the ground, for example, of protection of intellectual property and industrial secrets.

GDPR does not specify who should be the addressee of the demonstration of the security of processing personal data. Therefore, it may be claimed that it is enough to demonstrate the relevant properties to appropriate certification bodies which specialize in such verification services. The current framework of certification schemes includes strong state control of the member states over certification bodies. This may work well, as long as the authorities protect the interests of citizens. However, this is only the hypothetical ideal situation; recent history has shown that the reality has frequently differed. In particular, in dictatorships, any kind of certification by laboratories controlled by the state (either directly or informally) leads to a situation where verifica-

[9] See *Urbano Reviglio,* Towards a Right not to Be Deceived? An Interdisciplinary Analysis of Media Personalization in the Light of the GDPR. I3E Workshops 2019, p. 47–59.

tion does not protect citizens' rights, but rather serves the dictator and helps them make sure that the appropriate trapdoors are installed.

This problem was recognized more than a decade ago, in connection with attempts to design secure e-voting schemes. In that regard, **local verifiability** is requested. It means that the verification process cannot be entirely delegated to specialized certification bodies and that a voter has an opportunity to detect any kind of misconduct of election authorities, as well as of the electronic devices acting on behalf of the voter.

The challenging problem that arises due to anamorphic schemes is that a mere inspection of a block box device, based on its behavior, is insufficient. In other words, non-evasive local verification of black box devices is infeasible for a large number of cryptographic fundamental functionalities.

6. Defense

In the described circumstances, we should neither surrender the right to security of black box devices nor abandon the use of such devices. The situation is not hopeless from a technical point of view. From a legal point of view, it would be desirable to request for a relevant verification to be done by the user themselves, or at least by a third party. Such a third party should not only be independent of the device provider; it should be demonstrable that any kind of collusion is extremely unlikely. We note that a similar organizational framework has been created for NIST competitions for fundamental cryptographic primitives.

The technical countermeasures against threats of anamorphic schemes might be based on de-randomization: The lack of random ephemeral values makes it difficult to find space for hiding covert messages. Of course, this is not a simple issue: one cannot simply replace a random number generator with a deterministic unpredictable value (as is the case of EdDSA), for the sheer reason that it must be demonstrated that the final implementation is de-randomized.

Remote attestation – a pretty reliable and effective method – is not that easy in the case of black box devices. Great care must be taken to exclude the conversion of remote attestation to a global surveillance framework.

The most promising approach is to design schemes based on the following concepts:

1) **Watchdog:** a device controlled by the user, which is an interface between the black box device and the rest of the environment[10]. The watchdog may, for example, re-encrypt the ciphertexts, effectively destroying hidden channels.

[10] C.f. *S. Sherman*, *M. Chow*, *A. Russell*, *Q. Tang*, *M. Yung*, *Y. Zhao*, *H.S. Zhou*, Let a Non-barking Watchdog, p. 221–251.

2) **User secret components**: a device may be obliged to use secret values provided explicitly by the user. In this case, not only should the device be more immune to malicious implementations, but the user should also be given observable evidence that the user's components participate in the computation, as declared[11].
3) **Subversion evidence**: the last line of defense might be to design black box devices in such a way that, in the event of a leakage, undeniable evidence of a breach can be presented[12].

Although many solutions of this kind require some changes in cryptographic schemes, they could be local in some sense. For example, a verifiable signature creation device could be controlled by its owner (for this purpose, some additional output of the device might be necessary), while the final signature should have exactly the same shape as the original scheme.

Abstract

A trusted device is often a critical trust anchor in IT systems. Even if the rest of an IT system cannot be fully trusted, trusted devices should guarantee an acceptable level of overall security. To play this role, trusted devices should be secure by design. In this way, the effort is shifted from security analysis to inspection of compliance with a given "secure design". This has resulted in dependance on certification and audit schemes which are based on verifying characteristics included in a checklist.

If a certification process concerns devices that implement the concept of a black box device, then the examination process yields certificates concerning concrete inspected devices. However, it is extremely difficult to say whether the device presented for inspection and the device handed over to a user are identical.

In this paper, we discuss the threats of treacherous certification of black box cryptographic devices. We point to the results on anamorphic cryptographic schemes that may be applied by a supplier of a device, in order to create a stealth information leakage – including, in particular, the secret keys for the device. The attack has substantial practical potential and can be used against device users on a wide scale.

Streszczenie

Zaufane urządzenie jest często krytycznym punktem dla zapewnienia zaufania w systemach informatycznych. Nawet jeśli pozostałe komponenty systemu informatycznego nie są wiarygodne, takie zaufane urządzenia powinny gwarantować akceptowalny poziom ogólnego bezpieczeństwa. Aby móc sprostać tym wymaganiom, zaufane urządzenia powinny być bezpieczne w trybie „*by-design*". Pozwala to na przeniesienie wysiłku z analizy bezpieczeństwa urządzenia na kontrolę zgodności z przedłożoną bezpieczną architekturą urządzenia. Podejście to skutkuje uzależnieniem od systemów

[11] C.f. *L. Hanzlik*, *K. Kluczniak*, *M. Kutyłowski*, Controlled Randomness - A Defense Against Backdoors in Cryptographic Devices, MYCRYPT 2016, LNCS 10311, Springer Verlag, pp. 215–232.
[12] C.f. *B. Pfitzmann*, PhD Disseration, Fail-stop Signature Schemes, University of Hildesheim, 1995.

certyfikacji i audytu, opierających się na weryfikacji postulowanych cech urządzenia zawartych na liście kontrolnej.

Jeżeli proces certyfikacji dotyczy urządzeń realizujących koncepcję urządzenia typu czarnej skrzynki, wtedy wynik badań certyfikacyjnych dotyczy konkretnych urządzeń przedstawionych do kontroli. Jednak nie sposób jest stwierdzić, czy urządzenie przedstawione do kontroli i urządzenie przekazane użytkownikowi są identyczne.

W niniejszym rozdziale omawiamy zagrożenia związane ze zdradliwą certyfikacją urządzeń kryptograficznych typu *black box*. Wskazujemy na wyniki dotyczące anamorficznych schematów kryptograficznych, które mogą być zastosowane przez dostawcę urządzenia w celu stworzenia ukrytego wycieku informacji – w tym w szczególności tajnych kluczy zawartych w urządzeniu. Tego typu ataki mają znaczny potencjał praktyczny i mogą być stosowane przeciwko użytkownikom urządzeń na szeroką skalę.

Bogdan Fischer

Rozdział 2.
Metadane – otwartość danych i świadomość zagrożeń

1. Wstęp

Przyzwyczailiśmy się, że wykorzystujemy hiperłącza zrozumiałe dla człowieka, dzięki którym przemieszczamy się pomiędzy dokumentami, stronami www, a w tle funkcjonują agenci wykorzystujący metadane, które pozwalają na znacznie bardziej rozbudowane powiązanie komunikacyjne komputerów. Agent rozumiany jest jako oprogramowanie zdolne do monitorowania otoczenia i działania w sposób inteligentny/autonomiczny. Inteligentne czynności wykonywane bez udziału człowieka to w szczególności wyszukiwanie, łączenie, porządkowanie, dopasowanie na podstawie określonych kryteriów dla osiągnięcia założonych celów, np. zarządzania urządzeniami. W usprawnianiu komunikacji pomagają rozszerzenia oferujące standaryzujące modele. Dzięki ujednolicaniu struktur semantycznych[1] (z różnych domen) porozumiewanie komputerów bez udziału człowieka jest coraz bardziej złożone, przy jednoczesnym skracaniu czasu wykonywania zadań. Wzbogaca to wymianę informacji, wiedzę i oparte na niej możliwości. Informacja może być traktowana jak odrębny, subiektywny byt w zależności od tego jak jest przez użytkownika rozumiana i w jakim celu wykorzystywana. Aby była legalna, muszą być określone i przestrzegane prawa dostępu do informacji uwzględniające uregulowania systemów dostępowych w UE i państwach członkowskich. Szczególne znaczenie w tym obszarze ma upowszechnienie zasad otwartości danych i gwarancji kształtujących prawo do ponownego wykorzystywania informacji sektora publicznego. Sukcesywnie rozwijane w tym obszarze prawo miękkie

[1] Ale także syntaktycznych zawierających opis formatów danych, zasad kodowania, budowy struktur.

i twarde, w swojej ostatecznej, jak na razie, formie zostało ukształtowane przez dyrektywę 2019/1024 z 20.6.2019 r. w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego[2], która została implementowana w Polsce ustawą z 11.8.2021 r. o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego[3]. Od lat, wraz z ich faktyczną transformacją, rozwijane są gwarancje w ramach kategorii kształtującej prawo informacyjne w znaczeniu przedmiotowym, a więc jako kompleksu norm prawnych regulujących pozyskiwanie, generowanie, udostępnianie, wyłączanie, przetwarzanie czy rozpowszechnianie informacji, które dotyczą realizacji prawa do informacji, a także obowiązki administracji publicznej względem tych informacji, w tym procedur postępowania z nimi i ich ochrony. Nie jest to odseparowany obszar, ale rozbudowujący się system wspierany wolnościami i prawami, których realizacja wiąże się z dostępem do informacji czy też posługiwaniem się informacjami łącznie kształtującymi status informacyjny człowieka[4]. U podstaw otwierania danych i ponownego wykorzystywania leży idea efektywnej eksploatacji informacji, która jest możliwa nie tylko poprzez udostępnianie samej docelowej treści (obiektów), ale także metadanych, czy metadanych o metadanych. Pozwala to na bezpośrednie przemieszczanie się od rekordu (konkretnego kontekstu), metadanych do samego obiektu, umieszczonego w zasobach sieci. W przypadku metadanych o danych stanowiących informację sektora publicznego[5], należy zasadniczo traktować je również jako informacje sektora publicznego, które podlegają ponownemu wykorzystywaniu[6]. Dlatego m.in. archiwa są zobowiązane do udostępniania metadanych stanowiących opisy materiałów archiwalnych, biblioteki – opisy katalogowych obiektów bibliotecznych, a muzea – opisy muzealiów.

Prawna konstrukcja akcentująca znaczenie metadanych w powiązaniu z danymi jest odzwierciedleniem praktyki, w której dla uzyskania jednolitości pliku źródłowego oraz jego kopii, konieczne jest zapewnienie kompletności/zbieżności całego zestawu danych budujących określony plik, z uwzględnieniem metadanych. Z kolei publikowanie dobrej jakości metadanych[7], zdefiniowanych w powszechnie stosowanych standardach, pozwala na ich efektywną wymianę między systemami i wykorzystywanie zarówno z obiektem, do którego się odnoszą, jak i odrębnie od niego, co zależy od celu wykorzystania.

[2] Dz.Urz. UE L 172 z 2019 r., s. 56.

[3] Dz.U. z 2021 r. poz. 1641. Ustawa z 11.8.2021 r., zastąpiła ustawę z 25.2.2016 r. o ponownym wykorzystywaniu informacji sektora publicznego (t.j. Dz.U. z 2019 r. poz. 1446). Na temat historii rozwoju ponownego wykorzystywania informacji sektora publicznego zob. *B. Fischer*, *M. Sakowska-Baryła*, *A. Piskorz-Ryń*, *J. Wyporska-Frankiewicz*, Ustawa o otwartych danych, art. 1.

[4] Tamże; zob. *M. Sakowska-Baryła*, Ochrona danych osobowych, s. 44–45.

[5] Szerzej na temat informacji sektora publicznego i jej definicji *B. Fischer*, *M. Sakowska Baryła*, *A. Piskorz-Ryń*, *J. Wyporska-Frankiewicz*, Ustawa o otwartych danych, komentarz do art. 2 pkt 12.

[6] Odmienna sytuacja będzie miała miejsce, gdy metadane nadpisują dane prywatne.

[7] Przy czym jakość rozumiemy zgodnie z EN ISO 19101, jako całość cech produktu, które rzutują na jego zdolność do zaspokajania stwierdzonych i ukrytych potrzeb.

Podstawowe zastosowanie metadanych, uwzględnione w niniejszym rozdziale, to narzędzie tworzenia charakterystyk zasobów sieciowych (w szczególności informacji sektora publicznego).

2. Nie tylko *white hacking*

Celem rozdziału jest krótkie przenalizowanie pojęcia metadanych, prób jego definicji, potrzeb tworzenia metadanych surowych, ich agregacji, integracji i potrzeb współdziałania. Tworzenie danych i metadanych nie jest obecnie realizowane wyłącznie w procesach z udziałem ludzi, ale w dużej mierze, odbywa się na podstawie procedur zautomatyzowanych, bez bezpośredniego wpływu człowieka, albo półautomatycznie. W wymianie danych pochodzących z różnorodnych źródeł, przy ich udostępnianiu (przekazywaniu), wiele informacji zawartych jest w warstwie niewidocznej dla człowieka, stąd nie zawsze mamy świadomość jakie informacje mogą uzyskać inni poprzez wykorzystanie metadanych. Ze względu na powiązania, każdy znaleziony element otwiera drogę do kolejnego, co wiąże się z ryzykiem doprowadzenia do danych lub ich charakterystyki, które chcemy, aby były objęte tajemnicą. Udostępnianie zbioru danych z metadanymi umożliwia dokonanie różnorodnych ustaleń, jak np. pochodzenia, tj. „historii zbioru danych oraz jego cyklu życia od momentu zgromadzenia i pozyskania poprzez kompilację i doprowadzenie do jego aktualnej postaci”[8]. Do upowszechniania i ustandaryzowania metadanych, do czego nakłania poprzez *soft law* i *hard law* UE, należy dodać potrzebę uświadamiania znaczenia informacji, które mogą metadane ze sobą nieść. Nie ma wątpliwości, że są potrzebne oraz że warunkują rozwój i możliwości jednolitego rynku cyfrowego oraz że istotną kwestią jest stworzenie mechanizmu unifikującego różnorodne normy (także normy *de facto*) opisujące zasoby cyfrowe. I chociaż rozdział ma być przyczynkiem do dyskusji nie tylko o pozytywnych kontekstach zastosowania metadanych, ale i zagrożeniach, nie jest jego założeniem zbadanie metod wyszukiwania i analizy metadanych zawartych w upublicznionych dokumentach (ang. *metadata harvesting*), ani zasad działania i rozgraniczania *white*, *grey* i *black hackingu*[9] czy stanowiącego element

[8] Rozporządzenie Komisji (WE) Nr 1205/2008 z 3.12.2008 r. w sprawie wykonania dyrektywy 2007/2/WE Parlamentu Europejskiego i Rady w zakresie metadanych (Dz.Urz. UE L z 2008 r. Nr 326, s. 12). Załącznik Przepisy Wykonawcze Dotyczące Metadanych, cz. A Interpretacje pkt 1. Przepis ten wskazuje na działanie zgodne z normą EN ISO 19101.

[9] Zasadą *white hackingu*, określanego również jako etyczny hacking, jest wykorzystywanie narzędzi stosowanych przez hackerów dla pozyskania informacji cennych z punktu widzenia bezpieczeństwa, następnie wykorzystywanych dla przeciwdziałania ewentualnym atakom. *White hacker* to ekspert bezpieczeństwa sieci i komputerów, który specjalizuje się w zabezpieczeniu systemów informatycznych. Wykorzystuje wiele metodologii dla zapewnienia cyberbezpieczeństwa. Działa zgodnie z prawem, a jego wyłącznym celem jest wykazanie słabych punktów, które następnie są naprawiane i zabezpieczane. Uważany za kluczowy sposób na zapewnienie proaktywnej strategii bezpieczeństwa. Przeciwieństwem „etycznego hackera”, jest *black hat hacker*. Z kolei działania pośrednie określane są jako *grey*