

Wojciech Mincewicz

ORCID: 0000-0003-0460-9158

Metadane – cichy zabójca prywatności

SŁOWA KLUCZOWE:

metadane, anonimowość, bezpieczeństwo informacji, FOCA, nagłówki poczty elektronicznej

Wprowadzenie

W ciągu minuty na świecie wysłanych zostaje 156 mln e-maili¹, które zawierają tysiące terabajtów danych: zdjęć, dokumentów tekstowych i innych plików cyfrowych. Tylko promil użytkowników Internetu wie, że udostępniając plik tekstowy czy też zdjęcie dostarcza swojemu interlokutorowi informację na przykład o: dacie utworzenia pliku, nazwie komputera, imionach i nazwiskach twórców, informacje o współautorach, a w przypadku zdjęć na przykład informację o typie urządzenia rejestrującego obraz czy o miejscu, gdzie został uwieczniony obraz. Wszystkie te dane nazywane są metadanymi.

Pojęcie metadanych było tradycyjnie używane w katalogach bibliotecznych do lat 80. XX wieku, gdy biblioteki przekształciły swoje dane katalogowe w cyfrowe bazy danych. Metadane definiowane są w sposób opisowy jako dane o danych². Są to więc dane z przedrostkiem meta,

¹ Za. B. Marr, *Every Day? The Mind-Blowing Stats Everyone Should Read*, Forbes, 21.05.2018, <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#4ffe1f6360ba> (dostęp: 11.01.2019).

² Definicje taką odnajdujemy np. w: *Słownik encyklopedyczny informacji, języków i systemów informacyjno-naukowych*, B. Bojar (opr.), Warszawa 2002, s. 153; lub w: M.L. Zeng,

który swój źródłosłów bierze z języka greckiego i może oznaczać następstwo lub zmienność: poza, po, pod, łącznie, wśród, według, prze-³. Jako jedni z pierwszych temat danych o danych poruszali David Griffel i Stuart McIntosh, którzy w 1967 roku mówili o istnieniu tematycznych opisów danych oraz metajęzyku opisującym relacje danych⁴. W literaturze przedmiotu metadane definiowane są jako informacje strukturalne, które opisują, wyjaśniają, lokalizują lub w inny sposób ułatwiają pobieranie i wykorzystanie informacji⁵, dane połączone z opisywanymi obiektami, odciążające ich potencjalnych użytkowników od konieczności posiadania pełnej apriorycznej wiedzy o ich istnieniu lub charakterystykach⁶. Oryginalne spojrzenie na metadane przedstawia na swoim blogu Lorcan Dempsey, który pisze, że lubi myśleć o metadanych jako danych, które usuwają z użytkownika potrzebę posiadania pełnej wiedzy na temat istnienia lub cech potencjalnie interesujących środowisko⁷. Inna definicja dostępna w literaturze amerykańskiej podpowiada, że metadane oznaczają uporządkowane informacje o zasobie informacyjnym w dowolnym formacie⁸.

W literaturze polskiej kwestia metadanych zaczęła się pojawiać pod koniec XX wieku. Za prekursora badań uznawany jest Zdzisław Płoski, który metadane określa jako dane służące do opisu innych danych⁹. Inny z uznanych autorów Marek Nahotko rozszerza tę definicję i metadane określa jako: ustrukturyzowane, czytelne maszynowo dane, zawierające charakterystykę cyfrowych obiektów informacyjnych, które służą ich efek-

Metadata Basics, <http://marciazeng.slis.kent.edu/metadatabasics/cover.htm> (dostęp: 12.12.2018).

³ Za: *Słownik języka polskiego* PWN, M. Szymczak (red.), Warszawa 1995, s. 134. Zobacz także np.: *Słownik języka polskiego*, [b.a], Warszawa 1962, s. 577.

⁴ Zob. D. Griffel, S. McIntosh, *ADMINS – A Progres Raport*, MIT 1967, s. 27 i n., <https://dspace.mit.edu/bitstream/handle/1721.1/82974/09487802.pdf> (dostęp: 12.09.2018).

⁵ National Information Standards Organization, *Understanding Metadata*, Bethesda 2004, <https://web.archive.org/web/20141107022958/http://www.niso.org/publications/press/UnderstandingMetadata.pdf> [dostęp: 12.10.2018]. (Na temat NISO zob. przypis 20).

⁶ D. Campbell, *Dublin Core Metadata and the Australian Metaweb Project*, 10th National Library Technicians' Conference, Fremantle, 8–10.09.1999, <https://www.nla.gov.au/bla/staffpaper/dcampvell1.html> (dostęp: 15.10.2018).

⁷ L. Dempsey, *Registries: The Intelligence in Network*, 20.08.2006, <http://orweblog.oclc.org/Registries-the-intelligence-in-the-network/> (dostęp: 12.11.2018).

⁸ M. Foulonneau, J. Riley, *Metadata for Digital Resources. Implementation, Systems Design and Interoperability*, Oxford 2008, s. 3, za: P. Caplan, *Metadata Fundamentals for All Librarians*, Chicago 2003, s. 3.

⁹ Zob. Z. Płoski, *Informatyka: słownik encyklopedyczny*, Wrocław 1999, s. 56.

tywnemu oraz trafnemu wyszukiwaniu, szczególnie w wielkich zasobach informacji w Internecie¹⁰. W innych definicjach autorzy kładą z kolei nacisk na funkcję metadanych i twierdzą, że za ich pomocą opisywane są dokumenty cyfrowe, w szczególności dokumenty dostępne poprzez sieci komputerowe. Wśród przykładowych danych wymienia: pliki, katalogi, datę utworzenia czy informacje o modyfikacji¹¹. Nie jest to jednak zbiór zamknięty w zakresie funkcji, jakie spełniają metadane, albowiem wydaje się, że przede wszystkim służą one opisowi dokumentu, ale także ułatwiają wyszukiwanie danych, ukazują informacje o historii danych, ułatwiają wyszukiwanie określonych dokumentów, a także mogą ułatwić ocenę przydatności konkretnych danych.

W socjolekcie prawniczym za metadane uważane są logicznie powiązane z dokumentem elektronicznym usystematyzowane informacje opisujące ten dokument, które ułatwiają jego wyszukiwanie, kontrolę, zrozumienie i długotrwałe przechowywanie oraz zarządzanie¹².

Przytoczone rozumienia metadanych sprawiają, że dla dalszych rozważań niezbędne jest zaproponowanie definicji regulującej, która uściśli znaczenie kluczowego dla całego artykułu pojęcia i będzie definicją operacyjną dla studium. W tym celu posłużono się ustaleniami zawartymi w dokumencie *Standardy metadanych e-PL*, w którym przyjęto, że metadane to „wszelkie dane o dokumentach lub zbiorach dokumentów odnoszące się do ich treści, parametrów technicznych i fizycznych. Metadane mogą się odnosić do wszelkich dokumentów bez względu na sposób ich wytworzenia lub zapisu, w tym także do dokumentów elektronicznych. Metadane mogą określać istotne elementy takie jak tematykę dokumentów lub zbiorów dokumentów, osoby lub instytucji odpowiedzialnych za powstanie, czas wytworzenia, sposób zapisu, zasady dostępu itd.”¹³.

¹⁰ M. Nahotko, *Metadane – sposób na uporządkowanie Internetu*, Kraków 2004, s. 15. Bliżniaczą definicję odnajdujemy w naukach bibliotekoznawczych, gdzie metadane definiowane są jako ustrukturyzowane, zazwyczaj czytelne maszynowo dane, zawierające charakterystykę dokumentów służących ich efektywnemu oraz trafnemu wyszukiwaniu, zarządzaniu nimi i ich wartościowaniu. Zob. G. Czapnik, Z. Gruszka (red.), *Podręczny słownik bibliotekarza*, Warszawa 2011, s. 197.

¹¹ Zob. M. Niebrzydowska, R. Kotowicz, *Wstęp do informatyki śledczej*, „Przegląd Bezpieczeństwa Wewnętrznego” 2012, nr 6, s. 66.

¹² Zob. Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z 30 października 2006 r. w sprawie niezbędnych elementów struktury dokumentów elektronicznych (Dz.U. Nr 206, poz. 1517).

¹³ Naczelna Dyrekcja Archiwów Państwowych, *Standardy metadanych, e-PL, wersja 0.1*, Warszawa 2005, s. 7, <https://www.archiwa.gov.pl/images/docs/e-PL-0.1-2.pdf> (dostęp: 15.10.2018).

Definicja ta, chociaż jest wystarczająca dla zrozumienia problematyki, wymaga jednak doprecyzowania, albowiem w ocenie autora metadane należy rozumieć przede wszystkim jako ustrukturyzowane dane, zapisane w logiczny sposób, które w jakikolwiek sposób identyfikują i charakteryzują dokument elektroniczny. Jest to tzw. szerokie rozumienie problematyki, mające charakter prymarny wobec pozostałych przymiotów, które zależne będą od tego, z jakim typem metadanych mamy do czynienia. Wówczas mowa jest o zawężeniu definicji do konkretnej kategorii.

Typologie metadanych

W literaturze przedmiotu występuje co najmniej kilka propozycji klasyfikacji i systematyzacji metadanych. Aby usystematyzować pojęcie, należy rozróżnić co najmniej trzy podejścia do zagadnienia. Pierwsze jest charakterystyczne dla bibliotekoznawstwa i opiera się na wykorzystaniu metadanych do tworzenia opisów bibliograficznych materiałów bibliotecznych oraz ich wyszukiwaniu. Drugie podejście ma charakter bardziej informacyjny i pozwala na wykorzystanie metadanych dla zarządzania danymi, kładąc nacisk na ich funkcje administracyjne i zarządzające zasobami¹⁴. W tym przypadku podstawowym zadaniem metadanych jest dostarczenie uporządkowanej, logicznie spójnej dokumentacji, która opisuje konstrukcję, powstanie i sposób wykorzystania danych w określonym systemie komputerowym. Trzecie podejście, infobrokerskie, znajduje zastosowanie w codziennej pracy brokera informacji czy informatyce śledczej. Należy je sytuować pomiędzy podejściem bibliotekoznawczym i informacyjnym. W tym przypadku metadane zarówno będą służyły do tworzenia opisów plików, jak i znajdą zastosowanie w czynnościach operacyjnych, albowiem dane zdobyte w trakcie działań mogą stać się dowodem cyfrowym. Są to dane ukryte dla zwykłego użytkownika, zawarte w tworzonych w plikach, dokumentach elektronicznych. Mogą zostać pozyskane przy użyciu technik charakterystycznych dla białego wywiadu¹⁵. W takich cyfrowych „śmieciach” pozostawionych często nieświadomie przez użytkownika infobroker może odnaleźć interesujące go informacje, takie jak na przykład:

¹⁴ Por. M. Nahotko, *Metadane...*, s. 15.

¹⁵ Więcej na temat białego, szarego i czarnego wywiadu czytaj: J. Garlicki, D. Mider, W. Mincewicz, *Pozyskiwanie informacji w Internecie metodą Google Hacking – biały, szary czy czarny wywiad?*, „Przegląd Bezpieczeństwa Wewnętrznego” 2019, nr 20, s. 68-91.

- datę i czas utworzenia pliku;
- adres lub położenie geograficzne miejsca utworzenia pliku;
- imię i nazwisko, nazwę firmy, nazwę komputera lub adres IP;
- nazwy wszystkich współtwórców dokumentu lub dodane komentarze;
- typ użytego aparatu i jego ustawienia podczas robienia zdjęcia;
- typ użytego urządzenia rejestrującego audio lub wideo i jego ustawienia podczas nagrywania;
- informacje o marce lub modelu smartfona¹⁶.

Próba typologizacji problematyki metadanych również nie jest zadaniem łatwym. Na przykład Francis Patton Bretherton i Paul T. Singley rozróżniają dwie odrębne klasy metadanych: strukturalne i przewodnie. Metadane strukturalne opisują strukturę obiektów baz danych, takich jak tabele, kolumny i indeksy. Metadane przewodnie pomagają znaleźć określone przedmioty i są zwykle wyrażane jako zestaw słów kluczowych w języku naturalnym¹⁷. Rozszerzenie i ustrukturyzowanie tej typologizacji stanowi podejście różnicowania poprzez kryterium „typu danych”. Na tej podstawie można wyróżnić metadane dla danych pierwotnych i metadane przetwarzania danych. Dane pierwotne to wszystkie te, które zarządzane są przez źródła danych, zasoby danych, giełdy danych oraz aplikacje. Odpowiednie metadane zawierają informacje związane ze strukturą źródła danych i zasobów danych. Metadane przetwarzania to informacje związane z przetwarzaniem danych: dotyczące procesów ładowania i modyfikacji danych, ich analizy oraz administrowania¹⁸. Podziału metadanych można także dokonać z uwzględnieniem kryterium czasu, w którym zostały wytworzone. W tym przypadku możemy mówić o następujących danych o danych:

- zbieranych w czasie projektu;

¹⁶ Należy zdawać sobie sprawę, że metadane są zjawiskiem multidyscyplinarnym. Z uwagi na wymogi redakcyjne w artykule zaprezentowane zostały i zdefiniowane trzy podejścia do problematyki metadanych. W literaturze naukowej definiowane są one i opisywane m.in. w ekonomii, administracji czy też geolokalizacji. Zob. więcej odnośnie do np. administracji: M. Ganczar, *Informatyzacja administracji publicznej*, Warszawa 2009, s. 72–73; do ekonomii: S. Wrycza, *Informacja ekonomiczna. Podręcznik akademicki*, Warszawa 2010, s. 412 lub J. Zawila-Niedźwiecki, K. Rostek, A. Gąsiorkiewicz (red.), *Informacja gospodarcza 4*, Warszawa 2010, s. 289. Metadane znajdują także zastosowanie choćby w geolokalizacji – zob. L. Litwin, M. Rossa, *Metadane geoinformacyjne w INSPIRE i SDI*, Gliwice 2010.

¹⁷ Zob. F.P. Bretherton, P.T. Singley, *A User's View*, [w:] *Proceedings of the 7th International Working Conference on Scientific and Statistical Database Management*, Charlottesville 1994, s. 166–174.

¹⁸ M. Nahotko, *Metadane...*, s. 23.

- tworzonych w czasie powstawania systemu;
- tworzonych podczas uruchamiania systemu;
- statycznych;
- dynamicznych;
- krótkoterminowych;
- długoterminowych¹⁹.

Inne podejście do klasyfikacji metadanych odnajdujemy w zestawieniu przygotowanym przez National Information Standards Organization (NISO)²⁰. Wyróżniono tam trzy typy:

- metadane opisowe – służą do odnajdywania i identyfikacji kluczowych informacji, które umożliwiają lokalizację obiektu. Za przykład mogą służyć informacje o autorze, słowach kluczowych, wydawnictwie czy tytule;
- metadane strukturalne – opisują strukturę danego obiektu. W przypadku zbiorów bibliotecznych, które stanowiły *ad initium* dla zestawienia NISO, była to na przykład liczba rozdziałów, stron;
- metadane administracyjne – co odnosi się do informacji technicznych, w których zawarte są na przykład informacje o czasie i sposobie utworzenia pliku²¹.

Typologia proponowana w niniejszym artykule została przyjęta w sposób aprioryczny i stanowi punkt odniesienia do dalszych rozważań, w których dokonano jej egzemplifikacji na przykładzie dwóch najpopularniejszych formatów plików: 1. JPEG – Joint Photographic Experts Group²², charakterystycznego dla obrazów graficznych; 2. rozszerzenia

¹⁹ Typologia za: *Podstawowe funkcje metadanych*, <https://www.heuristic.pl/blog/internet/Co-to-sa-metadane;129.html> (dostęp: 11.11.2018).

²⁰ National Information Standards Organization to amerykańskie stowarzyszenie non-profit, które opracowuje normy w zakresie usług informacyjnych. NISO jest akredytowane przy amerykańskiej organizacji normalizacyjnej – American National Standards Institute. NISO opracowuje normy, raporty techniczne, zalecenia i inne dokumenty dla tradycyjnych procesów informacyjnych i nowych technologii, czyli w zakresie wszystkiego, co jest w jakikolwiek sposób związane z wyszukiwaniem, przetwarzaniem i ochroną informacji oraz metadanymi. Zostało założone w 1939 r. Informacja za oficjalną witryną: <https://groups.niso.org/home> (dostęp: 11.11.2018).

²¹ National Information Standards Organization, *Understanding...*; inne źródło, w którym przytoczono typologię: M.L. Zeng, *Metadata...*

²² Najpopularniejszym formatem zapisu metadanych plików graficznych są tagi Exchangeable Image File Format (EXIF). W przypadku aparatu cyfrowego podstawowymi metadanymi przechowywanymi na karcie pamięci są daty utworzenia pliku, ostatniej modyfikacji oraz ostatniego użycia. Sam aparat cyfrowy w momencie wykonania zdjęcia zapisuje w EXIF nazwę producenta urządzenia, model czy datę utworzenia pliku

.doc lub .docx, charakterystycznego dla plików tekstowych²³. Ekstrahowanie metadanych z plików i graficznych, i tekstowych możliwe jest za pomocą darmowych narzędzi dostępnych w przeglądarce Google, jak również ręcznie na komputerach osobistych.

Pracy nad artykułem autorowi przyświecały dwa cele. Pierwszy związany jest z uświadomieniem czytelnikowi potencjalnych zagrożeń i możliwości związanych z wykorzystaniem metadanych. Po drugie, chodziło o zawarcie takich elementów praktycznych, które pozwolą na samodzielne wykorzystanie technik defensywnych umożliwiających usunięcie metadanych z plików udostępnianych publicznie. Artykuł ma charakter autorskiego wyboru i opisuje najważniejsze zagadnienia, które umożliwiają zrozumienie problematyki opracowania.

Typologizacja metadanych w artykule odbywać się będzie według gradacyjnego ujęcia poszczególnych typów metadanych. Najpierw scharakteryzowane zostaną metadane opisowe, które są najszerszą kategorią identyfikującą autora pliku, następnie metadane strukturalne, które wprost opisują obiekt cyfrowy, a na końcu metadane administracyjne – zawierające dane techniczne pliku.

Metadane opisowe

Pierwszy typ metadanych wyróżnionych w toku powyższych rozważań to metadane opisowe (ang. *descriptive metadata*), które stanowią podstawę każdego dokumentu cyfrowego. Obejmują całą wiedzę, która gromadzona jest wokół obiektu. Mogą zawierać podstawowe informacje ewidencyjne, ale też szczegóły dotyczące historii czy przeznaczenia obiektu. Metadane

dla czasu ustawionego w aparacie cyfrowym. W metadanych pliku graficznego zlokalizowana jest także np. informacja o ustawieniach aparatu czy też pozycji GPS. Inne formaty zapisu metadanych plików graficznych to IPCT czy XMP. Zob. *Zanim wgrasz wakacyjne zdjęcie do sieci...*, Niebezpiecznik.pl, 2.07.2015, <https://niebezpiecznik.pl/post/zanim-wgrasz-wakacyjne-zdjecia-do-sieci/> 2 lipca 2015 (dostęp: 12.11.2018).

²³ W przypadku dokumentów tekstowych plik zawiera – oprócz samego tekstu wytworzonego przez użytkownika – dodatkowe różnorodne niejawne informacje dotyczące tego pliku. System operacyjny Windows przekazuje do metadanych informacje np. o datach: utworzenia, ostatniej modyfikacji i użycia pliku. Program MS Word wytwarza informacje np. o autorze, osobach, które edytowały plik, dacie ostatniego zapisania, ostatniego drukowania czy czasie edycji. Za: P. Wichań, *Informatyka śledcza*, post FB z 16.11.2015, <https://www.facebook.com/iswichran/posts/metadane-w-informatyce-czyli-dane-niejawne-ukryte-dla-zwyk%C5%82ego-u%C5%BCytkownika-zawar/1713057508925147/> (dostęp: 11.12.2018).

opisowe służą do opisu treści oraz elementu zbiorów, pomagają w łączeniu autorów ze zbiorem i zapewniają ważny kontekst dotyczący zasobu po jego wykryciu. Ten typ metadanych daje możliwość wyszukiwania, przeglądania, sortowania i filtrowania informacji. Metadane opisowe mogą zawierać takie elementy, jak tytuł, opis, dane autorów. Jest to najczęstszy typ metadanych, znany większości twórców treści cyfrowych. Służą one do identyfikowania oraz opisywania zbiorów. Odgrywają szczególną rolę przy sortowaniu zbiorów bibliotecznych, w których poszczególne pozycje identyfikowane są najczęściej poprzez autora lub tytuł.

Metadane opisowe plików z rozszerzeniem .doc i .docx tworzone są przez edytor Microsoft Word. Zgodnie z przytoczoną definicją metadane opisowe mają za zadanie w pierwszej kolejności pomóc wyszukać dany plik dzięki tytułowi, autorowi. Metadane opisowe w plikach Word w darmowy sposób wyodrębnić można na co najmniej dwa sposoby. Pierwszy – poprzez wybranie prawego przycisku myszy → właściwości → szczegóły. Wówczas odnajdziemy informacje, w których zawarty jest tytuł i temat naszego pliku, tagi, kategorie i komentarze w opisie, a także dane o autorze, osobach, które modyfikowały dokument, liczbie modyfikacji, a także potencjalnie instytucji, z którą związany jest autor dokumentu (rys. 1).

Rysunek 1. Metadane opisowe wyodrębnione z pliku o rozszerzenie .docx

Opis	
Tytuł	Infobrokering polityczny
Temat	Metadane, czyli dane o da...
Tagi	metadane.; Studia Politolog...
Kategorie	
Komentarze	
Pochodzenie	
Autorzy	mgr Wojciech Mincewicz
Ostatnio zapisany przez	Wojciech Mincewicz
Numer poprawki	4
Numer wersji	
Nazwa programu	Microsoft Office Word
Firma	Uniwersytet Warszawski
Menedżer	
Utworzenie zawartości	12.07.2018 14:28
Data ostatniego zapisania	03.08.2018 13:59
Ostatnio drukowany	25.07.2018 16:03
Całkowity czas edycji	09:41:00

Metadane opisowe pliku Word

Źródło: opracowanie własne.

Nie jest to jednak jedyny sposób wyodrębniania metadanych opisowych z dokumentu tekstowego. Możliwe jest także skorzystanie z dowol-

nego darmowego generatora online (rys. 2). Wówczas uzyskujemy podobny zestaw metadanych opisowych. Generatory dostarczają informacje w sposób mniej usystematyzowany, bardziej chaotyczny, jednakże zaletą jest to, że w przypadku dokumentów o rozszerzeniu .doc (starsza wersja pakietu Office) uzyskujemy więcej informacji.

Rysunek 2. Metadane opisowe wyodrębnione online z pliku o rozszerzenie .doc

File Type	DOC
File Type Extension	doc
Mime Type	application/msword
Title	Temat: Metadane, czyli dane o danych
Author	Wojciech Mincewicz
Template	Normal
Last Modified By	WOJTEK

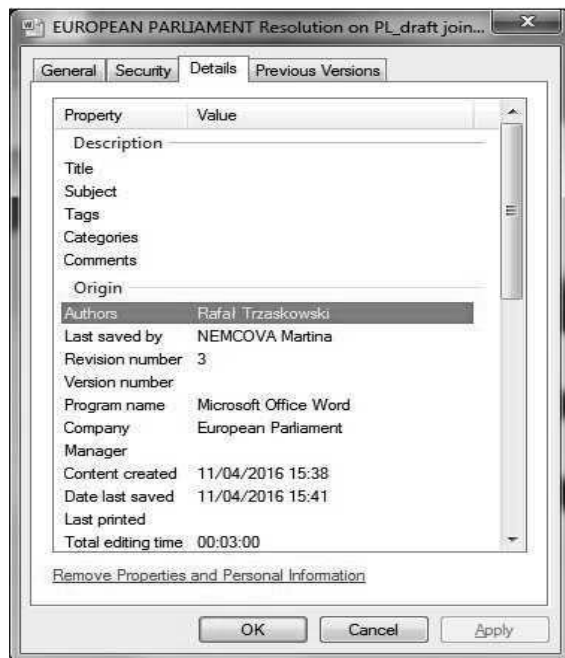
Źródło: opracowanie własne.

Metadane opisowe stanowią grupę danych, które zostały zaimplementowane do świata informatyki z bibliotekoznawstwa, gdzie spełniają porządkową funkcję i są jak najbardziej pożądane. W przypadku plików tekstowych, które przechowujemy na prywatnym dysku i które nie są nigdzie udostępniane, pełnią one podobną funkcję i są jak najbardziej właściwe. Sytuacja zmienia się jednak, gdy dokument, którego jesteśmy autorami, ma zostać udostępniony lub też umieszczony w Internecie. Wówczas warto zastanowić się, czy metadane opisowe nie mogą być źródłem problemów? Bolesnie przekonał się o tym Rafał Trzaskowski w kwietniu 2016 roku. Wówczas w Parlamencie Europejskim trwały prace nad rezolucją potępiającą zmiany w wymiarze sprawiedliwości²⁴. Po kilku dniach Tomasz Poręba, jeden z parlamentarzystów Prawa i Sprawiedliwości ujawnił na jednym z portali społecznościowych, że nad rezolucją, pracował Rafał Trzaskowski (rys. 3). Polityk Platformy Obywatelskiej, mimo początkowego dementi, przyznał się w jednym z programów publicystycznych do prac nad dokumentem, gdy jego interlokutor Adam Bielan powiedział o metadanych²⁵.

²⁴ Zob. np.: js.ika, *Europarlament przyjął rezolucję. Jakie będą kolejne kroki*, TVN24, PAP, 13.04.2016, <https://www.tvn24.pl/wiadomosci-z-kraju,3/europarlament-przyjal-rezolucje-ws-polski-jakie-beda-kolejne-kroki,635357.html> (dostęp: 11.11.2018).

²⁵ *Bielan: To są metadane, pan jest autorem. Trzaskowski: nie wstydzę się tego*, TVN24, 15.04.2016, <https://www.tvn24.pl/kropka-nad-i,3,m/kropka-nad-i-bielan-to-sa-metadane-pan-jest-autorem,636737.html> (dostęp: 11.12.2018).

Rysunek 3. Analiza metadanych projektu rezolucji przyjętej przez Parlament Europejski w kwietniu 2018 roku



Źródło: Twitter @TomaszPoreba, <https://twitter.com/tomaszporeba/status/720559877115011072> [dostęp: 24.11.2018].

Metadane opisowe plików z rozszerzeniem JPEG dostarczają przede wszystkim informacji o sprzęcie, na którym został uwieczniony obraz. Z metadanych opisowych zawartych w EXIF-ach odczytamy informację o samym aparacie, nazwie pliku (domyślnie zlokalizowana tam będzie data wykonania zdjęcia), obiektywie, ekspozycji (czyli ilości światła padającego na element światłoczuły), o lampie błyskowej czy skupieniu. Są więc one tworzone przez sam sprzęt, którym wykonujemy fotografię. W przypadku plików JPEG w ocenie autora o wiele bardziej przydatne do odczytania metadanych są narzędzia, które robią to w sposób automatyczny, a są dostępne w darmowej wersji. Ręczne odczytywanie metadanych (poprzez opcję właściwości → szczegóły) dostarcza dużo mniej informacji.

Rysunek 4. Metadane opisowe wyekstrahowane z tagów EXIF dla pliku z rozszerzeniem JPEG

Artist:	Picasa
Camera:	Nikon D60
Lens:	AF-S DX Zoom-Nikkor 18-55mm f/3.5-5.6G ED II Shot at 42 mm
Exposure:	Auto exposure, Not Defined, 1/60 sec, f/5.3, ISO 200
Flash:	Auto, Fired, Return detected
Focus:	AF-A, Mid-right, at <u>2.7m</u> , with a depth of field of about <u>86cm</u> , (from about <u>36cm</u> before the focus point to about <u>50cm</u> after) AF Area Mode: Dynamic Area (closest subject)
Date:	June 25, 2014 6:35:51PM (timezone not specified) (4 years, 5 months, 20 days, 5 hours, 8 minutes, 39 seconds ago, assuming image timezone of 1 hour ahead of GMT)
Time Zone Offset:	+01:00
File:	3,872 × 2,592 JPEG (10.0 megapixels) 3,623,255 bytes (3.5 megabytes)
Color Encoding:	WARNING: Color space tagged as sRGB, without an embedded color profile. Windows and Mac browsers and apps treat the colors randomly. Images for the web are most widely viewable when in the sRGB color space and with an embedded color profile. See my Introduction to Digital-Image Color Spaces for more information.

Źródło: opracowanie własne.

Metadane strukturalne

Metadane strukturalne (ang. *structural metadata*) to najbardziej niezrozumiana forma metadanych. Jest powszechnie ignorowana, nawet wśród osób, które pracują z metadanymi. Kiedy jest omawiana, może być mylona z innymi typami. Nawet osoby, które poprawnie rozumieją metadane strukturalne, nie zawsze doceniają ich pełny potencjał. Metadane strukturalne to dane dotyczące struktury treści. Pod pewnymi względami wcale nie jest to tajemnicze²⁶. Ułatwiają one nawigację i prezentację zasobów elektronicznych. Dostarczają informacji na temat wewnętrznej struktury zasobów, znaczników struktury, takich jak: nazwa strony tytułowej, informacje o spisie treści, tytułach rozdziałów i podrozdziałów. W przypadku obrazów metadane strukturalne identyfikować będą sam obraz, a nie jak wcześniej sprzęt, którym został on uwieczniony. Zawarte

²⁶ M. Andrews, *Structural Metadata. Key to Structured Content*, 11.10.2017, <https://storyneedle.com/structural-metadata-key-to-structured-content/> (dostęp: 12.12.2018).

tam będą dane między innymi o wymiarach obrazu czy jego rozdzielczości.

Metadane strukturalne plików z rozszerzeniem .doc i .docx powstają z automatu i są tworzone przez edytor Microsoft Word. Zgodnie z powyższą definicją dostarczają informacji o strukturze pliku. W przypadku plików z tym rozszerzeniem analizie podlega przede wszystkim zawartość pliku, to jest: liczba stron, statystyka wyrazów, liczba znaków, liczba wierszy, akapitów (rys. 5). Metadane strukturalne, podobnie jak opisowe, w plikach Word w darmowy sposób wyodrębnić można na co najmniej dwa sposoby ręcznie, jak również za pomocą generatorów online.

Rysunek 5. Przykładowe metadane strukturalne wyodrębnione z pliku .doc

Stan zawartości	
Typ zawartości	application/vnd.openxmlfor...
Strony	28
Statystyka wyrazów	4568
Liczba znaków	27408
Liczba wierszy	228
Liczba akapitów	63
Szablon	Normal
Skala	Nie
Linki brudne?	Nie
Język	

Źródło: opracowanie własne.

Generatory online, jak już sygnalizowano wcześniej, dostarczają informacje w sposób mniej ustrukturyzowany, ale są dokładniejsze, albowiem możemy tam odnaleźć na przykład informacje o spisie treści, czego nie zrobimy, pracując ręcznie (rys. 6).

Rysunek 6. Spis treści (dane zostały celowo usunięte ze względu na wrażliwe treści)

HYPERLINKS	
0	#_Toc516470996
1	#_Toc516470995
2	#_Toc516470994
3	#_Toc516470993
4	#_Toc516470992
5	#_Toc516470991
6	#_Toc516470990
7	#_Toc516470989
8	#_Toc516470988
9	#_Toc516470987
10	#_Toc516470983
11	#_Toc516470982
12	#_Toc516470981
13	#_Toc516470980
14	#_Toc516470979
15	#_Toc516470978
16	#_Toc516470977
17	#_Toc516470976
18	#_Toc516470975
19	#_Toc516470974
20	#_Toc516470973

Źródło: opracowanie własne.

Analiza metadanych strukturalnych fotografii zapisanej w formacie JPEG dostarcza brokerowi wiedzy na temat samego zdjęcia. Jest to informacja na temat oryginalnego rozszerzenia, wymiarów (szerokości, wysokości) czy rozdzielczości, a także szereg innych specjalistycznych informacji dotyczących samego zdjęcia związanych z kolorami czy jednostkami rozdzielczości.

Rysunek 7. Metadane strukturalne dla pliku zapisanego w formacie JPEG

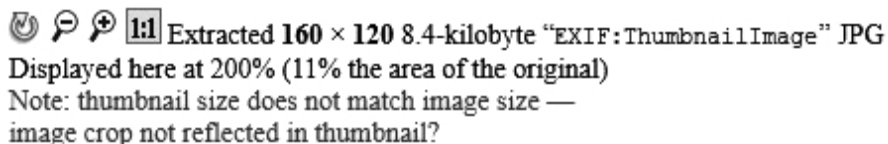
Composite
This block of data is computed based upon other items. Some of it may be wildly incorrect, especially if the image has been resized.

Aperture	5.30
Lens ID	AF-S DX Zoom-Nikkor 18-55mm f3.5-5.6G ED II
Lens Spec	18-55mm f3.5-5.6 G
Megapixels	10.0
Shutter Speed	1/60
Create Date	2014:06:25 18:35:51.70 4 years, 5 months, 19 days, 20 hours, 8 minutes, 39 seconds ago
Date/Time Original	2014:06:25 18:35:51.70 4 years, 5 months, 19 days, 20 hours, 8 minutes, 39 seconds ago
Modify Date	2014:09:23 22:22:16.70 4 years, 2 months, 21 days, 16 hours, 22 minutes, 14 seconds ago
Blue Balance	1.226563
Light Value	9.7
Red Balance	1.960938
Scale Factor To 35 mm Equivalent	1.5
Circle Of Confusion	0.020 mm
Depth Of Field	0.86 m (2.30 - 3.16 m)
Field Of View	31.4 deg (1.50 m)
Focal Length	42.0 mm (35 mm equivalent: 63.0 mm)
Hyperfocal Distance	16.62 m

Źródło: opracowanie własne.

Metadane strukturalne to także analiza miniatury zdjęcia, która pozwala stwierdzić, czy zdjęcie jest oryginalne, czy zostało poddane retuszowi, kadrowaniu lub innej obróbce. Dzięki informacji, która umieszczona jest na rysunku 8, można stwierdzić, że analizowane na potrzeby artykułu zdjęcie było edytowane, ponieważ kadrowanie obrazu nie odzwierciedla tego, co jest zawarte w miniaturze zdjęcia.

Rysunek 8. Metadane strukturalne miniatury pliku graficznego



Źródło: opracowanie własne.

Telefony i aparaty zapisują miniaturkę oryginalnego zdjęcia w tagach EXIF, ponieważ standard ten służy do wyświetlenia zdjęcia na widoku archiwum wykonanych zdjęć danego urządzenia. Dzieje się tak, gdyż aparaty czy telefony nie dysponują mocnymi procesorami, a miniaturka wyświetlana jest po prostu szybciej niż oryginalne zdjęcie wykonane w dużej rozdzielczości.

O miniaturkach zdjęć do końca życia będzie zapewne pamiętała amerykańska dziennikarka Catherine Schwartz. W 2003 roku opublikowała ona na swoim blogu kilka retuszowanych i przyciętych zdjęć. Jak się okazało, dzięki miniaturze fotografii, na niektórych z nich Schwartz była naga.

Metadane administracyjne

Metadane administracyjne (ang. *administrative metadata*) to dane techniczne, które powstają zazwyczaj automatycznie przy tworzeniu zasobu cyfrowego (zdjęcia lub pliku), a zawarte są w nich takie informacje, jak na przykład daty utworzenia, modyfikacji i użycia pliku, rozdzielczość, typ oraz model aparatu, przestrzeń kolorów, format pliku, kompresja, źródło światła, właściciel itp. W przypadku zdjęć cyfrowych – takie dane zapisane są bezpośrednio w pliku odwzorowania (na przykład w formacie EXIF) i nie powinno się ich usuwać – to cenne źródło informacji na temat procesu tworzenia zdjęcia i stanowią informację choćby o prawach autorskich.

Metadane administracyjne plików z rozszerzeniem .doc i .docx są generowane przez system operacyjny Windows, nie przez program. Zawierają podstawowe informacje o pliku, takie jak: rozmiar, data utworzenia, data ostatniej modyfikacji oraz ostatniego użycia. Stanowią grupę metadanych, które identyfikują nasz dokument. Ich wyodrębnienie, podobnie jak w przypadku metadanych opisowych i strukturalnych, możliwe jest na dwa sposoby – w sposób ręczny, a także za pomocą darmowych generatorów metadanych. Dane te w przypadku metadanych administracyjnych mają podobny kształt, jeżeli chodzi o strukturę (rys. 9 zawiera metadane wyodrębnione ręcznie, a rys. 10 – dane wyodrębnione za pomocą darmowego generatora).

Rysunek 9. Metadane administracyjne pliku tekstowego

Plik	
Rozmiar	367 KB
Data utworzenia	03.08.2018 12:11
Data modyfikacji	03.08.2018 13:59
Data dostępu	03.08.2018 13:59
Dostępność	

Źródło: opracowanie własne.

Rysunek 10. Metadane administracyjne pliku tekstowego, generator online

Software	Microsoft Office Word
Total Edit Time	1.7 hours
Last Printed	2018:06:11 07:08:00
Create Date	2018:06:11 04:55:00
Modify Date	2018:06:11 07:09:00
Pages	148
Words	41172
Characters	247034
Security	None
Company	Hewlett-Packard
Lines	2058
Paragraphs	575

Źródło: opracowanie własne.

Metadane administracyjne plików z rozszerzeniem JPEG stanowią zdecydowanie największą i najciekawszą z perspektywy infobrokera grupę danych, jakie przechowywane są w EXIF-ach. Zlokalizowane są tam

bowiem informacje identyfikujące sam plik, czyli w pierwszej kolejności o dacie powstania i modyfikacji, co umożliwia sprawdzenie, czy plik był w jakikolwiek sposób zmieniany.

Rysunek 11. Metadane administracyjne dla pliku graficznego

Composite

This block of data is computed based upon other items. Some of it may be wildly incorrect, especially if the image has been resized.

Aperture	5.30
Lens ID	AF-S DX Zoom-Nikkor 18-55mm f3.5-5.6G ED II
Lens Spec	18-55mm f3.5-5.6 G
Megapixels	10.0
Shutter Speed	1/60
Create Date	2014:06:25 18:35:51.70 4 years, 5 months, 19 days, 20 hours, 8 minutes, 39 seconds ago
Date/Time Original	2014:06:25 18:35:51.70 4 years, 5 months, 19 days, 20 hours, 8 minutes, 39 seconds ago
Modify Date	2014:09:23 22:22:16.70 4 years, 2 months, 21 days, 16 hours, 22 minutes, 14 seconds ago
Blue Balance	1.226563
Light Value	9.7
Red Balance	1.960938
Scale Factor To 35 mm Equivalent	1.5
Circle Of Confusion	0.020 mm
Depth Of Field	0.86 m (2.30 - 3.16 m)
Field Of View	31.4 deg (1.50 m)
Focal Length	42.0 mm (35 mm equivalent: 63.0 mm)
Hyperfocal Distance	16.62 m

Źródło: opracowanie własne.

W metadanych administracyjnych plików JPEG przy odrobinie szczęścia i niewadze autora zdjęcia odnajdziemy także koordynaty GPS. Większość urządzeń mobilnych automatycznie generuje lokalizację, co niekiedy bywa zgubne. Bolesnie przekonał się o tym Higinio O. Ochoa III, jeden z członków grupy Anonymous o nazwie Cabin Cr3w, który wykrał i upublicznił w Internecie dane setki policjantów. Funkcjonariusze bez problemu byli w stanie zlokalizować hakera, a stało się to dzięki koordynatom GPS zdjęcia, które pozostały na stronie internetowej policji²⁷.

Inny osobliwy przykład tego samego typu stanowi John McAfee, programista i twórca firmy McAfee Associates, który popadł w konflikt

²⁷ Za. S. Hang, *The Life of an Ex-Hacker Who Is Now Banned from Using the Internet*, 25.04.2015, <https://gizmodo.com/the-life-of-an-ex-hacker-who-is-now-banned-from-using-t-1700074684> (dostęp: 12.12.2018). (Higinio O. Ochoa został skazany przez sąd na zakaz korzystania z Internetu – nie może dotknąć niczego, co jest do niego podłączone. Gdy chce obejrzeć film, musi to czynić w towarzystwie żony).

z prawem. W 2012 roku w Belize został posądzony o zabójstwo swojego sąsiada. Tuż po przesłuchaniu w tej sprawie przez tamtejszy sąd uciekł. W ręce organów ścigania wpadł w Gwatemali, gdzie się ukrywał. Do swojej kryjówki zaprosił dziennikarzy magazynu „Vice”, którzy dokumentowali całość spotkania fotografiami, które opublikowali, nie usuwając przed tym koordynatów GPS²⁸.

Rysunek 12. Metadane administracyjne zawierające koordynaty GPS

Basic Image Information	
Target image: https://christianheilmann.com/wp-content/uploads/2014/10/me.jpg	
Camera:	Lge Nexus 5
Lens:	1.2 mm
Exposure:	1/40 sec, f/2.9, ISO 102
Flash:	none
Date:	October 19, 2014 4:06:20PM (timezone not specified) (5 years, 24 days, 20 hours, 34 minutes, 54 seconds ago, assuming image timezone of 1 hour ahead of GMT)
Location:	Latitude/longitude: 59° 19' 6.8" North, 18° 3' 35.5" East (59.318554, 18.059870) Map via embedded coordinates at: Google, Yahoo, WikiMapia, OpenStreetMap, Bing (also see the Google Maps pane below) Altitude: 0 meters (0 feet) Camera Pointing: East-southeast Timezone guess from earthtools.org: 1 hour ahead of GMT
File:	1,280 × 960 JPEG (1.2 megapixels) 425,979 bytes (416 kilobytes)
Color Encoding:	WARNING: Color space tagged as sRGB, without an embedded color profile. Windows and Mac browsers and apps treat the colors randomly. Images for the web are most widely viewable when in the sRGB color space and with an embedded color profile. See my Introduction to Digital-Image Color Spaces for more information.

Źródło: opracowanie własne.

Sposoby usuwania metadanych

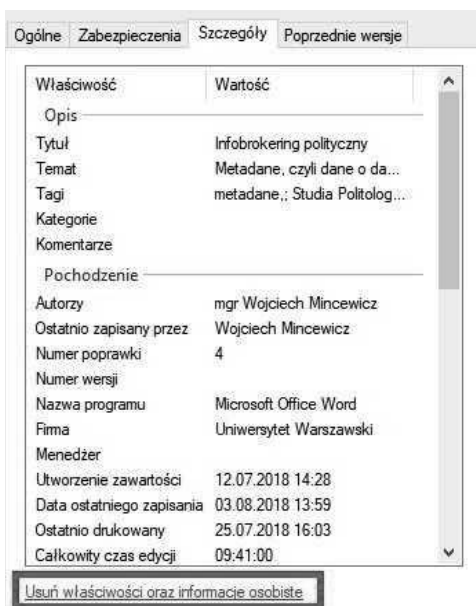
Znając już typy metadanych, można udzielić odpowiedzi na pytanie, które metadane plików tekstowych i graficznych usuwać i jak to robić tak, aby udostępniając dokument zyskać pewność, że czynimy to bez szkody dla naszej prywatności. Warto podkreślić, że ta część rozważań stanowi autorskie przemyślenia i nie powinna być w żadnym stopniu uznana za rozstrzygającą. Nie ma także jednej unikalnej metody, która zagwarantuje

²⁸ C.G. Weissman, *The Insane Life of Former Fugitive and Eccentric Cybersecurity Legend John McAfee*, 23.06.2015, <https://www.businessinsider.com/the-insane-life-of-john-mcafee-2015-7?IR=T> (dostęp: 13.12.2018).

użytkownikom całkowite bezpieczeństwo. Znając *case study* przytoczone w artykule oraz typologie przedstawioną powyżej, każdy sam powinien odpowiedzieć na pytanie, które dane usunąć i kiedy. Warto pamiętać, jaką rolę odgrywają poszczególne typy i jakie dane są tam przechowywane.

Metadane opisowe w pierwszej kolejności odgrywają rolę pomocniczą. Mają nam – użytkownikom ułatwić wyszukiwanie treści i spełniają pożyteczną funkcję, gdy są przechowywane w plikach na własnym komputerze. Przytoczony w artykule przykład dobitnie pokazuje, że mogą być jednak źródłem problemów, dlatego w ocenie autora usuwanie tych danych powinno być złotą zasadą postępowania z metadanymi opisowymi, gdy udostępniamy dokument. Jak to uczynić? W przypadku dokumentów z rozszerzeniem .docx metadane opisowe można usunąć ręcznie, wybierając usuń właściwości → informacje osobiste (zob. rys. 13). Wówczas w kolejnym oknie usuwamy: tytuł, temat, tagi oraz dane autora. Dane te możemy także usuwać ręcznie, poprzez ich edytowanie.

Rysunek 13. Ręczne usuwanie metadanych z pliku tekstowego



Źródło: opracowanie własne.

Inna metoda to mechaniczne usunięcie danych ukrytych w dokumencie, dzięki darmowemu programowi na przykład Doc Scrubber (rys. 14). Jest to jedno z narzędzi firmy Javacool Software, znanej między innymi ze Spyware Blastera, służące do przeglądania i usuwania ukrytych danych

zaszytych. Jest dedykowane szczególnie dla starszych wersji plików z rozszerzeniem .doc, których nie możemy usunąć ręcznie na naszym komputerze osobistym²⁹. Bliźniacze rozwiązania w zakresie mechanicznego usuwania metadanych zapisanych w tagach EXIF jest Easy Exif Delete.

Rysunek 14. Widok z programu DOC Scrubber



Źródło: opracowanie własne.

W przypadku dokumentów tekstowych, jak również zdjęć dobrym i skutecznym sposobem jest zastanowienie się, czy wszystkie formaty zapisu przechowują metadane. W artykule zostały omówione najpopularniejsze formaty zapisów plików tekstowych i graficznych, to jest .doc/.docx i JPEG. Można na przykład, zamiast rozpowszechniania dokumentu programu Word, przekonwertować dokument na format .rtf, .txt lub do klasycznego pdf. W przypadku obrazów zamiast formatu JPEG stosować format PNG. Formaty te przechowują małe ilości meta znaczników lub nie przechowują ich wcale.

²⁹ Leszek Ignatowicz sugeruje, że najlepszym sposobem na usunięcie metadanych z plików tekstowych zapisanych w formacie .docx jest użycie inspektora dokumentów. Zob. L. Ignatowicz, *Cyfrowe ślady mówią. Poradnik ochrony prywatności*, Warszawa 2015, s. 7. Jest to kwestia co najmniej dyskusyjna, albowiem w ten sposób usunięte zostaną tylko metadane opisowe pliku. Bez zmian pozostaną metadane administracyjne, które dostarczają informacji na temat daty powstania i modyfikacji dokumentu.

Rysunek 15. Analiza pliku graficznego zapisanego w formacie PNG³⁰**Basic Image Information**

Target file: hostessy.png

File:	960 × 640 PNG 920,804 bytes (0.88 megabytes)
Color Encoding:	WARNING: No color-space metadata and no embedded color profile: Windows and Mac web browsers treat colors randomly. Images for the web are most widely viewable when in the sRGB color space and with an embedded color profile. See my Introduction to Digital-Image Color Spaces for more information.

Źródło: opracowanie własne.

Analiza skrajnych przypadków przytoczonych w artykule uzmysławia, że przed udostępnieniem pliku tekstowego lub cyfrowego warto usunąć z niego metadane opisowe, które pozwalają identyfikować twórcę dokumentu. Wraz z rozwojem Internetu zaczęły powstawać pierwsze programy, które służyły do masowego odnajdywania i analizowania metadanych zawartych w upubliczniczonych dokumentach (ang. *metadata harvesting*). Metodyka ich działania nakierowana jest na pozyskiwanie danych ukrytych pozostawionych na witrynach internetowych dokumentów. Za pomocą odpowiedniego zapytania skierowanego do internetowej wyszukiwarki lokalizowane są dokumenty opublikowane przez interesującą nas organizację. Po ich zlokalizowaniu program analizuje pozostawione przez nieświadomych użytkowników metadane.

Jednym z najpopularniejszych programów przeznaczonych do *metadata harvesting* jest Fingerprinting Organizations with Collected Archives (FOCA). Program został stworzony przez firmę ElevenPaths w celu wyszukiwania, pobierania i analizowania dokumentów dla pozyskania cyfrowych informacji pozostawionych przez użytkowników świadomie bądź nie w zasobach Internetu³¹. Mechanizm działania FOCA jest bar-

³⁰ Format PNG przechowuje tylko szcztkowe metadane strukturalne dla plików graficznych (rozmiar), nie zachowuje natomiast metadanych opisowych i administracyjnych.

³¹ Pozyskiwanie i wyodrębnianie metadanych nie jest jedynym zastosowaniem programu FOCA, ponieważ za jego pomocą można również m.in.: analizować budowę witryny, identyfikować Domain Name System, wyszukiwać kopie zapasowe strony, wykrywać katalogi lub luki w zabezpieczeniach. Jednak większość z tych funkcji dostępna jest w pełnej, płatnej wersji programu. Darmowa wersja dostępna obecnie (styczeń 2019) w wersji 3.4 umożliwia skorzystanie z podstawowych funkcji programu, wśród których znajduje się analiza metadanych dokumentów. Pełen opis wszystkich funkcji programu dostępny na stronie producenta: <https://www.elevenpaths.com/labstools/foca/index.html> (dostęp: 1.01.2019).

dzo prosty: za pomocą odpowiedniego zapytania kierowanego do wyszukiwarki – Google, Bing i Exalead, odnajdujemy dokumenty opublikowane przez interesującą nas organizację i analizuje zawarte w nich metadane.

Rysunek 16. Widok programu FOCA po utworzeniu nowego projektu i wybraniu docelowej domeny



The screenshot shows the FOCA application window. At the top center is the logo, which consists of the letters 'FOCA' in a bold, blocky font, with a stylized, grey, frog-like creature with large eyes and a small mouth positioned between the 'O' and 'C'. Below the logo, the interface is organized into several rows of input fields and controls:

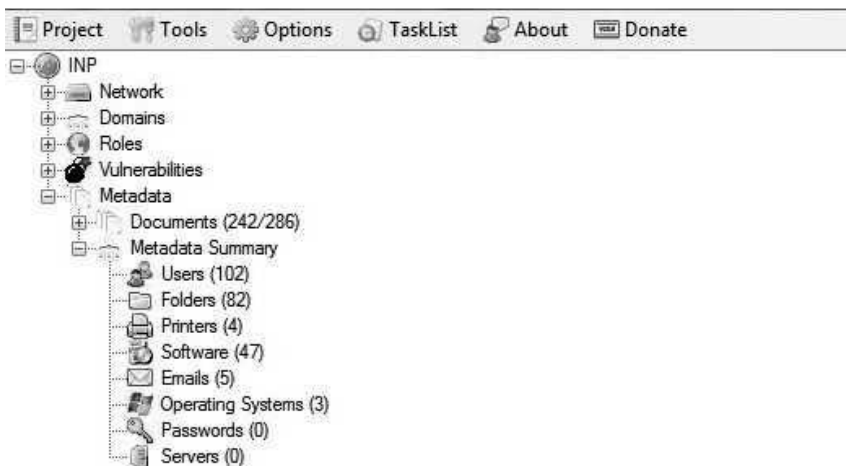
- Project name:** A text box containing 'INP'.
- Domain website:** A text box containing 'inp.uw.edu.pl'.
- Alternative domains:** An empty text box with a small up/down arrow icon on the right side.
- Folder where save documents:** A text box containing 'C:\Users\wmincewicz\Desktop\metac', followed by a folder icon button.
- Project date:** A text box containing '12.12.2018 18:14:53'.
- Project notes:** A text box containing 'Projekt testowy'.
- Autosave project each:** A numeric spinner box set to '0', followed by the word 'minutes'.

At the bottom of the window, there are two buttons: 'Update' on the left and 'Cancel' on the right.

Źródło: opracowanie własne.

Po utworzeniu nowego projektu i wybraniu docelowej domeny (rys. 16) możemy od razu rozpocząć poszukiwanie publicznie dostępnych dokumentów. Program FOCA bez problemu odnajdzie dokumenty umieszczone na testowanej domenie. Po zakończeniu wyszukiwania i wybraniu opcji Download All można pobrać wszystkie odnalezione pliki, a dzięki Extract ALL Metadata / Analyze All Metadata otrzymujemy zbiór metadanych, które uzyskał dla nas FOCA.

Rysunek 17. Widok programu FOCA po wyodrębnieniu metadanych z plików odnalezionych na analizowanej domenie



Źródło: opracowanie własne.

Finalnym efektem pracy programu jest pozyskanie przez infobrokera w pierwszej kolejności metadanych opisowych plików umieszczonych na analizowanej domenie (102 pliki miały nazwy użytkowników). Inne pozyskane dane to informacja o ścieżce zapisu plików, o sprzęcie oraz oprogramowaniu, którym posługują się autorzy plików. Program pozyskał także pięć adresów e-mail oraz informacje o systemach operacyjnych. W skrajnie ekstremalnych przypadkach za pomocą narzędzi do zmechanizowanego pozyskiwania metadanych pentesterzy³² pozyskują informacje o hasłach pozostawionych przez nieświadomych użytkowników. Metadane administracyjne oraz strukturalne plików, które zostały pozyskane z testowanej domeny, pozyskujemy w sposób „klasyczny” opisany w pierwszej części artykułu.

Jednym z najpopularniejszych sposobów przekazywania plików tekstowych i graficznych jest przesyłanie ich za pomocą poczty e-mail. Czy jeśli przed wysłaniem pliku w załączniku poczty elektronicznej usuniemy metadane opisowe, które identyfikują nasz dokument, to czy nasza przesyłka jest całkowicie pozbawiona meta znaczników? Absolutnie nie. Nawet przy zachowaniu wszelkich środków cyberhigieny szereg danych ukrytych zawiera sam nasz e-mail, a właściwie sam jego nagłówek. Dane

³² Pentester – osoba przeprowadzająca kontrolowany atak na jakiś system teleinformatyczny w celu oceny bieżącego stanu zabezpieczenia tego systemu; www.nowewyrazy.uw.edu.pl/haslo/pentester.html?pdf=1 (dostęp: 19.06.2019).

te mogą służyć między innymi potwierdzeniu autentyczności wiadomości otrzymanej z nieznanego źródła czy zlokalizowaniu nadawcy.

Pierwszą grupę metadanych, które zawiera nagłówek e-mail, są metadane opisowe pozwalające zidentyfikować nadawcę wiadomości. Meta znaczniki tam zawarte pozwalają stwierdzić, kto jest nadawcą e-maila, kiedy została napisana wiadomość, jaki jest jej ID oraz adres serwera gmail, gdzie zapisana jest wiadomość oraz kto jest jej adresatem.

Rysunek 18. Metadane opisowe nagłówka e-mail³³

```
MIME-Version: 1.0
From: Daniel Mider <d.mider@uw.edu.pl>
Date: Tue, 6 Nov 2018 16:49:09 +0100
Message-ID: <CAH7c=1eRUSuWj1AkJZdHGGkwa618HZY31R87Jk-s2gKtpY@mail.gmail.com>
Subject: Szkolenie FGI - materiały
To: '
        , Wojciech Mincewicz <w.mincewicz@student.uw.edu.pl>
Content-Type: multipart/mixed; boundary="00000000000f8ec80857a00f1ea"
```

Źródło: opracowanie własne.

Kolejną grupę metadanych zawartych w nagłówku e-mail stanowią te, które należy zaklasyfikować jako metadane administracyjne. W pierwszej kolejności będą to informacje, które identyfikują czas powstania e-maila: kiedy został wysłany przez autora, kiedy znalazł się na serwerze DNS i kiedy został dostarczony do adresata(ów) wiadomości. Czas ten w nagłówku e-maila zapisuje się według czasu pacyficznego standardowego (ang. *Philippine Standard Time*, PST)³⁴. Widoczne na rysunku 19 pozostałe składowe metadanych administracyjnych nagłówka e-mail – *Authenticated Received Chain*, uwierzytelniają kolejne serwery DNS, przez które przesyłana jest wiadomość e-mail. ARC składa się z trzech kluczowych komponentów: *Authentication-Results*; *Seal*; *Message-Signature*. Z perspektywy infobrokera najistotniejszą rolę spełnia pierwszy z nich, odpowiadający za weryfikację autentyczności otrzymanej wiadomości. Jeżeli e-mail został pozytywnie zweryfikowany przez uwierzytelnianie DMARC (ang. *Domain-based Message Authentication, Reporting, and Conformance*), zostanie on przekazany do odczytu. Podczas sprawdzania autentyczności DMARC korzysta z rekordów SPF (ang. *Sender Policy Framework*) i DKIM (ang. *DomainKeys Identified Mail*). Wiadomość, która nie przejdzie weryfikacji SPF lub DKIM, uruchomi zasady DMARC

³³ Pozostali adresaci e-maila zostali w sposób celowy zanonimizowani.

³⁴ Autor w pracy nad artykułem posługuje się adresem e-mail w aplikacji gmail.com, której serwery zlokalizowane są w Stanach Zjednoczonych Ameryki, dlatego przy analizie nagłówka e-mail otrzymujemy informację o czasie strefy czasowej PST.

i najprawdopodobniej nie dotrze do adresata, ponieważ zostanie uznana za spam lub niebezpieczną przesyłkę³⁵.

Rysunek 19. Metadane administracyjne nagłówka e-mail

```

Delivered-To: w.mincewicz@student.uw.edu.pl
Received: by 2002:a17:90a:c482:0:0:0:0 with SMTP id j2-v6csp4181879pjt;
  Tue, 6 Nov 2018 07:49:33 -0800 (PST)
X-Received: by 2002:a17:902:a516:: with SMTP id s22-v6mr5214202plq.255.1541519366055;
  Tue, 06 Nov 2018 07:49:26 -0800 (PST)
ARC-Seal: i=1; a=rsha-sha256; b=1541519366; cv=none;
  d=google.com; s=arc-20160816;
  b=cy4c8pRam5xw4kduJzjd13+t11wohNpAtD09DUc5+LwL+Y7Kc3D8gYfRcsLlvh0PN
  lhgVUQSEyppU3MnfWXz+eYRnnicWZdScforXhscb7XvE/dtyuVHBvF6FRpWyUk7VhZGZu
  EWZwJ77wMVZnFcM/U31rbRkzJxqMb7zqHbfutEVYmU1Em+Qpx6fnixAawyZlqUmK2eH
  MWguUm7gw6o2XURUFAIAjMgAcLtt2VvG7DpxY2aCFPNFVnqUHQYyAA+lqce2Zn8UASdiB
  XpDMLzvgTfa/JAMG0aWUuKbLukNRmpa+BS4JWSovJGN3WFxNjcmkpLA2OimQ8Q7KgLP
  iZJA==
ARC-Message-Signature: i=1; a=rsha-sha256; c=relaxed/relaxed; d=google.com; s=arc-20160816;
  h=to:subject:message-id:date:from:mime-version:dkim-signature;
  bh=s9doghbBV+wP4/G+/aEryI2QB+QSPKm5Gkwk49wNcm4=;
  b=s/zqUG6BivwtTU1LcSwZJ5BgcQl;4RWtGU3kav2LOSCH6pnO/pnlS3LgnWLX175x2n
  6skEoCGGzr+Nt0E15ZVAQpNrVcXOfiV9Ttpo2WjaCRr7F8CBFYgaVKdpKqkUp7FuiWh
  uc+J8PkmpBURMGLStukObyP4ckrdEm9VrLf1sIISiaSQiFvWzL0e9w9pbNcnBv3Q0
  YG12HW+Slm/5tdffffWnzuz9ANW593w664MDEgVcXI8c4jSA+jQzyR5qRAXYFddmFG3z
  PizDSe6/MzEdXpWBCvzuC4CXgyCEPFuN601+guHGDF0UepZEgYiRax9wyyqxDS6MB8Z
  6lKA==
ARC-Authentication-Results: i=1; mx.google.com;
  dkim=pass header.i=@uw.edu.pl.20150623.gappssmtp.com header.s=20150623 header.b=qllLftpF;
  spf=pass (google.com: domain of d.mider@uw.edu.pl designates 209.85.220.41 as permitted sender)
  smtp.mailfrom=d.mider@uw.edu.pl
Return-Path: <d.mider@uw.edu.pl>
Received: from mail-sor-f41.google.com (mail-sor-f41.google.com. [209.85.220.41])
  by mx.google.com with SMTPS id b17-v6sor4488211pls.11.2018.11.06.07.49.25
  for <w.mincewicz@student.uw.edu.pl>
  (Google Transport Security);
  Tue, 06 Nov 2018 07:49:25 -0800 (PST)
Received-SPF: pass (google.com: domain of d.mider@uw.edu.pl designates 209.85.220.41 as permitted sender) client-
  ip=209.85.220.41;

```

Źródło: opracowanie własne.

Ostatni typ metadanych, który możemy zidentyfikować w nagłówku e-maila, to grupa metadanych strukturalnych, najtrudniej definiowalnych. W przypadku e-maila, który stanowi *case study* dla artykułu, metadanymi strukturalnymi będą dane o załączniku do e-maila. W innych przypadkach za metadane strukturalne uznawane będą na przykład informacje o czonce, które odpowiedzialne są za formatowanie stopki e-maila.

³⁵ Więcej na ten temat: G. Colburn, *How to Explain Authenticated Received Chain (ARC) in Plain English*, 6.11.2015, <https://blog.returnpath.com/how-to-explain-authenticated-received-chain-arc-in-plain-english-2/> (dostęp: 2.01.2019).

Rysunek 20. Metadane strukturalne nagłówka zawierające informacje o czcionkach w stopce e-maila

```

<html xmlns:v=3D"urn:schemas-microsoft-com:vml" xmlns:o=3D"urn:schemas-microsoft-com:office:office" xmlns:w=3D"urn:schemas-microsoft-com:office:word" xmlns:m=3D"http://schemas.microsoft.com/office/2004/12/omml" xmlns=3D"http://www.w3.org/TR/REC-html40">
<head>
<meta http-equiv=3D"Content-Type" content=3D"text/html; charset=3Diso-8859-2">
<meta name=3D"Generator" content=3D"Microsoft Word 15 (filtered medium)">
<style><!--
/* Font Definitions */
@font-face
=09{font-family:"Cambria Math";
=09panose-1:2 4 5 3 5 4 6 3 2 4;}
@font-face
=09{font-family:Calibri;
=09panose-1:2 15 5 2 2 2 4 3 2 4;}
/* Style Definitions */
p.MsoNormal, li.MsoNormal, div.MsoNormal
=09{margin:0cm;
=09margin-bottom:.0001pt;
=09font-size:11.0pt;
=09font-family:"Calibri", sans-serif;
=09mso-fareast-language:EN-US;}
a:link, span.MsoHyperlink
=09{mso-style-priority:99;

```

Źródło: opracowanie własne.

Podsumowanie

Rozwój społeczeństwa informacyjnego, które stanowi następstwo społeczeństwa przemysłowego, możliwy jest dzięki rozwojowi techniki. Ten z kolei napędzany jest przez rozwój Internetu. Świat w realiach XXI wieku staje się globalną wioską, w której kolejne obszary życia społeczeństw przenoszą się do wirtualnej przestrzeni. Transfer ten oraz rozwój technologii i coraz łatwiejszy do nich dostęp są na pewno szansą na dalszą prosperitę, ale stanowi także zagrożenie dla naszej prywatności i anonimowości. Organizacje zajmujące się bezpieczeństwem, korporacje czy inne podmioty zabiegają o pozyskanie jak najszerszej gamy informacji o każdym z nas. Motywacje ich działania nie zawsze są w pełni zgodne z wyobrażeniami użytkowników Internetu. Informacjami, które sami produkujemy, wytwarzając kolejne pliki tekstowe lub graficzne, są metadane, czyli ustrukturyzowane dane zapisane w logiczny sposób, które w różnorodny sposób identyfikują i charakteryzują dokument elektroniczny. Mogą w nich być zawarte informacje identyfikujące dokument, jak również samego autora pliku.

Większość metadanych nie jest sama w sobie szkodliwa, pełnią one wręcz w jakimś stopniu pożyteczną funkcję. Przykładem mogą być metadane opisowe, które służą do opisu treści oraz elementu dokumentu. Ten

typ metadanych daje możliwość wyszukiwania pliku w ogromnych zbiorach danych poprzez tytuł lub dane autora. Służą więc one do identyfikowania oraz opisywania zbiorów, zatem w wypadku, gdy nikt poza osobami uprawnionymi nie ma do nich dostępu, są pożądane i pełnią pomocniczą funkcję. Problem pojawia się jednak, gdy plik tekstowy zostanie udostępniony lub zamieszczony w ogólnodostępnym Internecie. Wówczas warto zastanowić się, czy chcemy, aby każdy z użytkowników Internetu był w stanie identyfikować nas jako twórców danego dokumentu. Aby nie było to możliwe, warto znać podstawowe techniki defensywne, które mają służyć usunięciu metadanych z pliku. W trosce o swoją anonimowość oraz prywatność warto z którejs z nich skorzystać, szczególnie w sytuacji, gdy wiemy, że nasz plik będzie dostępny publicznie i może zostać pozyskany przez każdego internautę.

STRESZCZENIE

W artykule dokonano pogłębionej refleksji nad zagadnieniem metadanych, czyli danych, które definiują lub opisują inne dane. W warstwie teoretycznej wyodrębnione zostały trzy typy metadanych: opisowe, strukturalne i administracyjne. Metadane opisowe służą do odnajdywania i identyfikacji kluczowych informacji, które umożliwiają lokalizację obiektu. Metadane strukturalne opisują strukturę wewnętrzną danego obiektu, metadane administracyjne odnoszą się do informacji technicznych, zawarte są tam informacje na przykład o czasie i sposobie utworzenia pliku. Celem publikacji jest dostarczenie wiedzy teoretycznej, jak również praktycznej. W drugiej części artykułu dokonana została egzemplifikacja pojęcia na przykładzie plików graficznych i tekstowych oraz wskazane zostały proste techniki samoobrony, które umożliwiają usunięcie metadanych przed udostępnieniem pliku. Uzupełnienie tekstu stanowi analiza możliwości pozyskiwania meta informacji za pomocą programu Fingerprinting Organizations with Collected Archives (FOCA), który służy do zmechanizowanego pozyskiwania metadanych, a także refleksja nad tym, jakie metadane zawiera nagłówek e-maila.

Wojciech Mincewicz

METADATA – A SILENT PRIVACY KILLER

The article has a deeper reflection on the issue of metadata, that is, data which are defined or describe other data. The theoretical layer extracted three types of

metadata: descriptive, structural, and administrative. Descriptive metadata is used to find and identify key information that allows the location of an object. Structured metadata describes the internal structure of the object, but administrative metadata refers to the technical information, where information is provided for example about the time and how the file was created. The purpose of the publication is to provide theoretical knowledge as well as practical. The second part of the article depicts the concepts of graphic and text files, and simple self-defense techniques are indicated, which allow you to remove metadata before sharing the file. The supplementing of article is: analysis the ability to extract meta information by Fingerprinting Organizations with Collected Archives (FOCA), which is used to mechanizedly extract metadata reflection on what the metadata includes the email header.

KEY WORDS: *metadata, anonymity, information security, FOCA, e-mail header*

Bibliografia

- Andrews M., *Structural Metadata. Key to Structured Content*, 11.10.2017, <https://storyneedle.com/structural-metadata-key-to-structured-content/> (dostęp: 14.12.2018).
- Bretherton F.P., Singley P.T., *A User's View* [w:] *Proceedings of the 7th International Working Conference on Scientific and Statistical Database Management*, Charlottesville 1994.
- Campbell D., *Dublin Core Metadata and the Australian Metaweb Project*, 10th National Library Technicians' Conference, Fremantle, 8–10.09.1999, <https://www.nla.gov.au/bla/staffpaper/dcampvell1.html> (dostęp: 15.10.2018).
- Caplan P., *Metadata Fundamentals for All Librarians*, Chicago 2003.
- Colburn G., *How to Explain Authenticated Received Chain (ARC) in Plain English*, 6.11.2015, <https://blog.returnpath.com/how-to-explain-authenticated-received-chain-arc-in-plain-english-2/> (dostęp: 15.01.2019).
- Dempsey L., *Registries: The Intelligence in Network*, 20.08.2006, <http://orweblog.oclc.org/Registries-the-intelligence-in-the-network/> (dostęp: 11.11.2018).
- Foulonneau M., Riley J., *Metadata for Digital Resources. Implementation, Systems Design and Interoperability*, Oxford 2008.
- Garlicki J., Mider D., Mincewicz W., *Pozyskiwanie informacji w Internecie metodą Google Hacking – biały, szary czy czarny wywiad?*, „Przegląd Bezpieczeństwa Wewnętrznego” 2019, nr 20.
- Griffel D., McIntosh S., *ADMINS – A Progres Raport*, MIT 1967, <https://dspace.mit.edu/bitstream/handle/1721.1/82974/09487802.pdf> (dostęp: 20.09.2018).
- Hang S., *The Life of an Ex-Hacker Who Is Now Banned from Using the Internet*, 25.04.2015, <https://gizmodo.com/the-life-of-an-ex-hacker-who-is-now-banned-from-using-t-1700074684> (dostęp: 14.12.2018).
- Ignatowicz L., *Cyfrowe ślady mówią. Poradnik ochrony prywatności*, Warszawa 2015.
- Litwin L., Rossa M., *Metadane geoinformacyjne w INSPIRE i SDI*, Gliwice 2010.
- Marr. B., *Every Day? The Mind-Blowing Stats Everyone Should Read*, Forbes, 21.05.2018, <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create->

every-day-the-mind-blowing-stats-everyone-should-read/#4ffe1f6360ba (dostęp: 15.01.2019).

Nahotko M., *Metadane – sposób na uporządkowanie Internetu*, Kraków 2004.

Niebrzydowska M., Kotowicz R., *Wstęp do informatyki śledczej*, „Przegląd Bezpieczeństwa Wewnętrznego” 2012, nr 6.

Weissman C.G., *The Insane Life of Former Fugitive and Eccentric Cybersecurity Legend John McAfee*, 23.06.2015, <https://www.businessinsider.com/the-insane-life-of-john-mcafee-2015-7?IR=T> (dostęp: 15.12.2018).

Zeng, M.L. *Metadata Basics*, <http://marciazeng.slis.kent.edu/metadatabasics/cover.htm> (dostęp: 14.11.2018).