Marek Salamonowicz
Uniwersytet Warmińsko-Mazurski w Olsztynie
ORCID: 0000-0001-6934-2154
m.salamonowicz@uwm.edu.pl

https://doi.org/10.26881/gsp.2022.2.12

Otwarty dostęp do danych badawczych w ramach repozytorium instytucjonalnego a ochrona danych osobowych i prawa do prywatności

Wprowadzenie

Celem artykułu jest ustalenie optymalnych mechanizmów zapobiegania naruszeniom prawa do prywatności, jakie mogą być zastosowane w związku z otwartym dostępem do danych badawczych w ramach repozytoriów instytucjonalnych[1]. Analizie poddano szereg praktyk rekomendowanych w szczególności przez unijne lub krajowe organy administracji publicznej, a także zawartych w opracowaniach powstałych na zlecenie lub finansowanych przez instytucje publiczne[2]. Wywód oparto w szczególności na analizie dogmatyczno-prawnej unormowań dyrektywy Parlamentu Europejskiego i Rady (UE) 2019/1024 z dnia 20 czerwca 2019 r. w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego[3]. Wyrazem implementacji postanowień wymienionego aktu do krajowego porządku prawnego jest uchwalenie i wejście w życie ustawy z dnia 11 sierpnia 2021 r. o otwartych danych i ponownym

[1] Publikacja została napisana w wyniku odbywania przez autora stażu na Uniwersytecie w Walencji oraz w Urzędzie Unii Europejskiej ds. Własności Intelektualnej, współfinansowanego przez Unię Europejską w ramach Europejskiego Funduszu Społecznego (Program Operacyjnego Wiedza Edukacja Rozwój), zrealizowanego w projekcie Program Rozwojowy Uniwersytetu Warmińsko-Mazurskiego w Olsztynie (POWR.03.05.00-00-Z310/17).

[2] Zalecenie Komisji Europejskiej (UE) 2018/790 z 25 kwietnia 2018 r. w sprawie dostępu do informacji naukowej oraz jej ochrony, powstałej w wyniku badań naukowych finansowanych ze środków publicznych (Dz. Urz. UE z L 134, s. 12); L. Naudts, F. van den Boom, T. Tagarev *et al.*, *Best Practices, Blueprints and Policy Guidelines for Open Access to Scientific Information*, Leuven 2013, s. 8 https://cordis.europa.eu/docs/projects/cnect/1/325101/080/deliverables/001-OSLD83BestPracticesBluePrintsandPolicyGuidelinesforOpenAccesstoScientificInformation.pdf [dostęp: 9.01.2022]; Ministerstwo Nauki I Szkolnictwa Wyższego, *Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce*, Warszawa 2015, s. 4, https://www.gov.pl/documents/1068557/1069061/20180413_Kierunki_rozwoju_OD_wersja_ostateczna.pdf [dostęp: 9.01.2022]; Narodowe Centrum Nauki należy do „Coalition S", w ramach której wypracowano zasady otwartego dostępu oraz instrumenty jego wdrożenia; zob. Plan S – Making full and immediate Open Access a reality, https://www.coalition-s.org/ [dostęp: 9.01.2022].

[3] Dz. Urz. UE L 172, s. 56; dalej: dyrektywa 2019/1024.

wykorzystywaniu informacji sektora publicznego[4]. Odwołano się także do odpowiednich unormowań rozporządzenia Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE[5].

Polityki poszczególnych instytucji publicznych finansujących badania naukowe, w tym Europejskiej Rady Badań (*European Research Council*) w programie Horyzont Europa[6] czy Narodowego Centrum Nauki, nakłada na beneficjentów obowiązek udostępniania danych badawczych na zasadach FAIR (*findable*, *accessible*, *interoperable*, *reusable*), czyli możliwych do znalezienia, dostępnych, interoperacyjnych i wielokrotnego użytku. Wynika to z zasady, że dane badawcze powinny być „tak otwarte, jak to możliwe, tak zamknięte, jak to konieczne"[7]. Zasada ta jest również obecna w zaleceniu Komisji Europejskiej (UE) 2018/790 z dnia 25 kwietnia 2018 r. w sprawie dostępu do informacji naukowej oraz jej ochrony, powstałej w wyniku badań naukowych finansowanych ze środków publicznych[8]. Wspomniany dokument wskazuje powody, które mogłyby uzasadniać ograniczenia w otwartym dostępie do danych badawczych, w szczególności prywatność, tajemnicę handlową, bezpieczeństwo narodowe, uzasadnione interesy handlowe oraz do praw własności intelektualnej osób trzecich. Przedstawione zasady zostały także wyrażone w dyrektywie 2019/1024 (art. 10 ust. 1). Ustawodawca polski poprzez treść art. 22 ustawy z 2021 r. o otwartych danych nakłada na podmioty nauki i szkolnictwa wyższego[9] obowiązek publicznego udostępnienia w systemie teleinformatycznym tychże podmiotów, w szczególności w repozytorium instytucjonalnym lub tematycznym, danych badawczych wytworzonych lub zgromadzonych w ramach działalności naukowej, finansowanej ze środków publicznych, i które są już publicznie udostępniane przez te podmioty w systemie teleinformatycznym[10]. Takie dane podlegają ponadto bezpłatnie ponownemu wykorzystywaniu. Podmiot zobowiązany został przy tym upoważniony do określenia warunków ponownego wykorzystania danych.

---

4 Dz. U. poz. 1641; dalej: ustawa z 2021 r. o otwartych danych; por. W. Miller, *Nowa ustawa o ponownym wykorzystywaniu informacji sektora publicznego*, Temidium 2021, nr 3, s. 40.

5 Dz. Urz. UE L 119, s. 1; dalej: RODO; zob. D. Ossowska-Salamonowicz, *Ochrona danych osobowych w działalności dziennikarskiej*, Olsztyn 2015, s. 11.

6 *European Research Council (ERC) Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020*, Bruksela 2017, s. 3 i n. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf [dostęp: 9.01.2022].

7 Art. 10 dyrektywy 2019/1024, przypis 3.

8 Dz. Urz. UE L 134, s. 12.

9 Podmioty te zostały określone w art. 7 ust. 1 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (tekst jedn.: Dz. U. z 2021 r., poz. 478 ze zm.). Są nimi m.in. uczelnie, Państwowa Akademia Nauk i jej instytuty, instytuty badawcze, także międzynarodowe, Państwowa Akademia Umiejętności oraz Centrum Łukasiewicz.

10 W. Miller, *Nowa ustawa*…, s. 44; B. Fischer, *Autorskoprawne konteksty ponownego wykorzystania danych badawczych* [w:] *Sto lat polskiego prawa handlowego. Księga jubileuszowa dedykowana Profesorowi Andrzejowi Kidybie*, red. M. Dumkiewicz, K. Kopaczyńska-Pieczniak, J. Szczotka, t. 2, Warszawa 2020, s. 553.

Możliwe jest także udostępnienie danych badawczych bez szczególnych warunków ponownego wykorzystania danych. Wnosić należy, że przykładowo uczelnie publiczne jako podmioty zobowiązane nie mają obowiązku udostępniać wszystkich danych badawczych, jakie wytworzyły, korzystając ze środków publicznych, i zapewnić ich ponowne wykorzystanie, ale tylko tych, które już publicznie udostępniają.

Uzasadnieniem dążenia do zapewnienia otwartego dostępu do danych badawczych jest fakt, że taki dostęp przynosi szereg korzyści. Wśród korzyści płynących z otwartego udostępniania danych badawczych można wskazać te, które służą interesowi publicznemu, jak w szczególności bardziej odpowiedzialne i efektywne wykorzystanie środków publicznych, w tym ograniczanie ryzyka nieracjonalnie powtarzanych badań i utraty danych badawczych poprzez ich zachowanie. Zwiększenie użyteczności istniejących zbiorów danych badawczych rzutuje na lepsze możliwości niezależnej weryfikacji wyników badań naukowych oraz zwiększenie poziomu wykrywalności plagiatów, błędów i nierzetelności w prowadzonych badaniach naukowych[11]. Należy także wskazać na osobiste i zawodowe korzyści płynące z otwartego udostępniania danych badawczych, takie jak: możliwość większego uznania naukowego i możliwości współpracy badawczej, walidacja opublikowanych wyników badań, a w rezultacie ich większa wiarygodność. W szerszej perspektywie, otwarty dostęp do danych badawczych zapewnia ogromne korzyści w zakresie postępu technicznego i gospodarczego, podniesienia poziomu życia i rozwoju społeczno-kulturalnego.

W ciągu ostatnich kilku lat można zaobserwować wzrost liczby różnych typów repozytoriów w państwach członkowskich Unii Europejskiej, w tym w Polsce. Powstają rządowe repozytoria danych, repozytoria instytucjonalne, dziedzinowe (tematyczne) czy dedykowane określonemu projektowi czy programowi. W niniejszym materiale skupiono uwagę na repozytoriach instytucjonalnych danych badawczych[12].

Główne problemy na styku otwartego dostępu do danych badawczych z prawem do prywatności

W związku z funkcjonowaniem repozytoriów instytucjonalnych danych, wartości i zasady wynikające z ochrony prywatności, w tym danych osobowych, zderzają się z dążeniem do zapewnienia otwartego dostępu do danych badawczych i ich jak największej użyteczności. Z jednej strony występuje tendencja do zapewnienia przejrzystości i otwartości danych oraz możliwości ich wtórnego wykorzystania, a z drugiej – z wymogami prawnymi dotyczącymi ochrony danych osobowych. Powstaje

[11] P. Arzberger, P. Schroeder, A. Beaulieu, *Promoting access to public research data for scientific, economic, and social development*, "Data Science Journal" 2004, nr 3, s. 135 i n.; P.H. Dawson, S.Q. Yang, *Institutional repositories, open access and copyright: what are the practices and implications?*, "Science & Technology Libraries" 2016, vol. 35(4), s. 279 i n.

[12] N. Bhuva, E. Luca, *Issues and challenges in researchers' adoption of open access and institutional repositories: a contextual study of a university repository*, "Information Research: an international electronic journal" 2017, vol. 22(4), s. 6.

w szczególności problem ochrony prawa do prywatności w kontekście zaawansowanych technologii przetwarzania danych, w tym technologii *block-chain*, technologiami opartymi na sztucznej inteligencji, odwoływanie się do triangulacji danych czy obliczeń kwantowych[13]. Wymienione nowe technologie usprawniają korzystanie z danych badawczych, ale także mogą zwiększać ryzyko identyfikacji osób i naruszeń prawa do prywatności[14]. W wielu sytuacjach rozwiązanie tych kwestii sprowadza się do uzyskania właściwej równowagi pomiędzy skuteczną anonimizacją danych badawczych a zachowaniem ich użyteczności w możliwie największym stopniu. Stąd do głównych problemów na styku otwartego dostępu do danych badawczych z prawem do prywatności należy zaliczyć stworzenie procedur umieszczania danych badawczych w repozytorium, w tym wymogi dotyczące zgody na wykorzystanie danych osobowych w pracach badawczych uwzględniające kwestię udostępniania wyników badań naukowych. Niektóre ze środków ochrony prywatności uczestników badań powinny mieć umocowanie w umowie licencyjnej pomiędzy administratorem danych a podmiotem prowadzącym repozytorium instytucjonalne oraz umowach zawieranych z użytkownikami danych repozytorium instytucjonalnego. Ponadto, należy wskazać na zagadnienie anonimizacji i pseudonimizacji danych, ryzyko reidentyfikacji w kontekście stopnia wrażliwości danych i określenia poziomu dostępności i dopuszczalności transferu danych badawczych.

Zgodnie z art. 2 pkt 9 dyrektywy 2019/1024, „dane badawcze" oznaczają dokumenty w formie cyfrowej, inne niż publikacje naukowe, które są gromadzone lub produkowane w ramach działalności badawczo-naukowej i są wykorzystywane jako dowody w procesie badawczym bądź też są powszechnie akceptowane w środowisku naukowym jako konieczne do weryfikacji poprawności ustaleń i wyników badań. Przytoczona definicja legalna danych badawczych odwołuje się zatem do formy utrwalenia danych, wskazując na „dokumenty cyfrowe". Ponadto, wskazuje na element funkcji, jakie spełniają dane badawcze, tj. dowodzenie w procesie badawczym lub umożliwianie weryfikacji poprawności wyników i ustaleń. Nie wzbudza wątpliwości to, że dane badawcze powstają w ramach działalności badawczo-naukowej. Wykluczono z zakresu pojęcia danych badawczych publikacje naukowe. Wydaje się to uzasadnione, gdyż w zakresie otwartego dostępu do publikacji naukowych obowiązują szczególne uregulowania. Dane badawcze mogą obejmować informacje o wynikach eksperymentu, badania lub pomiaru, w tym metadane i dane dotyczące przetwarzania. Przykładowo dane takie obejmują statystyki, wyniki obserwacji, ankiet, nagrania wywiadów i obrazy. Definicja legalna danych badawczych z art. 2 pkt 2 ustawy z 2021 r. o otwartych danych również

[13] J. Atik, V. Jeutner, *Quantum computing and computational law*, "Law, Innovation and Technology" 2021, vol. 12(2), s. 303; C. Kuner, F. Cate, O. Lynskey *et al.*, *Blockchain versus data protection*, "International Data Privacy Law" 2018, vol. 8(2), s. 103.

[14] Por. wyrok TS z 7.05.2009 r. w sprawie C-553/07 Rijkeboer, ECLI:EU:C:2009:293, pkt 47 oraz wyrok TS z 8.04.2014 r. w połączonych sprawach C-293/12 i C-594/12 *Digital Rights Ireland and Others*, ECLI:EU:C:2014:238, pkt 53.

odwołuje się do kryterium funkcji tych danych w procesie dowodzenia lub weryfikacji wyników badań oraz formy elektronicznej[15].

Z kolei art. 4 pkt 1 rozporządzenia 2016/679, wskazuje na definicję danych osobowych, które oznaczają informacje o zidentyfikowanej lub możliwej do zidentyfikowania osobie fizycznej („osobie, której dane dotyczą"). Osoba fizyczna możliwa do zidentyfikowania to zgodnie z rozporządzeniem 2016/679 osoba, którą można bezpośrednio lub pośrednio zidentyfikować. Identyfikacja może w szczególności nastąpić na podstawie identyfikatora takiego jak imię i nazwisko, numer identyfikacyjny, dane o lokalizacji, identyfikator internetowy lub jeden bądź kilka szczególnych czynników określających fizyczną, fizjologiczną, genetyczną, psychiczną, ekonomiczną, kulturową lub społeczną tożsamość osoby fizycznej. Należy także wskazać, że dane badawcze obejmują nierzadko tzw. dane wrażliwe, jak np. dane kliniczne, które wymagają szczególnej ochrony, o czym szerzej w dalszej części niniejszego artykułu[16].

Jeżeli zatem dane osobowe, czyli wszelkie informacje dotyczące zidentyfikowanych lub możliwych do zidentyfikowania osób fizycznych, są udostępniane publicznie on-line za pośrednictwem otwartego repozytorium instytucjonalnego, należy uwzględnić przepisy dotyczące prywatności i ochrony danych osobowych. Natomiast dane, które nie odnoszą się do możliwych do zidentyfikowania osób, takie jak dane zbiorcze lub statystyczne, nie są zasadniczo danymi osobowymi. Dlatego też dane w pełni zanonimizowane nie są objęte zakresem przepisów o ochronie danych. Pod pojęciem „anonimizacja" prawodawca unijny rozumie proces zmiany danych w informacje anonimowe, które nie odnoszą się do zidentyfikowanej lub możliwej do zidentyfikowania osoby fizycznej, lub dane osobowe zanonimizowane w taki sposób, że identyfikacja osoby, której dane dotyczą, nie jest lub już nie jest możliwa[17]. Należy jednak podkreślić, że dane pseudonimizowane nadal umożliwiają identyfikację osoby, której dane dotyczą, dzięki połączeniu pseudonimu (np. kodu klucza, numeru kodu) z dodatkowymi identyfikatorami. Wówczas zgodnie z art. 2 pkt 5 RODO, są one tak przetworzone, by nie można ich było już przypisać konkretnej osobie, której dane dotyczą, bez użycia dodatkowych informacji, pod warunkiem że takie dodatkowe informacje są przechowywane osobno i są objęte środkami technicznymi i organizacyjnymi uniemożliwiającymi ich przypisanie zidentyfikowanej lub możliwej do zidentyfikowania osobie fizycznej[18].

---

15 R. Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym. Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, Warszawa 2021, s. 359.

16 P. Fajgielski, *Komentarz do rozporządzenia nr 2016/679 w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych)* [w:] *Ogólne rozporządzenie o ochronie danych. Ustawa o ochronie danych osobowych. Komentarz*, red. P. Fajgielski, LEX/el., komentarz do art. 9 pkt 5.

17 D. Lubasz, W. Chomiczewski, M. Czerniawski *et al.* [w:] *RODO. Ogólne rozporządzenie o ochronie danych. Komentarz*, red. E. Bielak-Jomaa, LEX/el., komentarz do art. 4, pkt 9.

18 *Ibidem*, komentarz do art. 4 pkt 5, pkt 1; M. Więckowska, *Anonimizacja i pseudonimizacja to główne narzędzia pracy administratorów danych, podwyższające bezpieczeństwo danych oraz osób, których dane są udostępniane i przetwarzane*, „ABI Expert" 2016, nr 1; M. Hintze, K. El Emam, *Comparing the benefits of pseudonymisation and anonymisation under the GDPR*, "Journal of Data Protection & Privacy" 2018, vol. 2(2), s. 145 i n.

Zasadnie należy zalecać, aby zbiory danych w otwartym dostępie nie zawierały danych osobowych. Powinny one być raczej anonimizowane, z reguły przez naukowca, który uzyskał dane badawcze, jest ich dysponentem i wnosi o ich udostępnienie poprzez repozytorium instytucjonalne. Jeżeli dane badawcze obejmowały wcześniej dane osobowe, w tym szczególnie wrażliwe, za dobrą praktykę należy uznać stosowanie wyraźnego zakazu ponownego identyfikowania danych zanonimizowanych lub pseudoanonimizowanych wobec licencjobiorców[19].

Niektóre dane badawcze są zagregowane, uzyskuje się je poprzez łączenie danych indywidualnych. Z perspektywy prowadzonych tu rozważań należy zauważyć, że udostępnianie danych zagregowanych jest bezpieczniejsze. Nie prowadzi ono bowiem do ujawnienia informacji o osobach fizycznych. Należy zaznaczyć, że zagregowane dane badawcze nie zawsze pozwalają na pełną odtwarzalność wyników i są mniej korzystne z punktu widzenia wykorzystania w dalszych badaniach naukowych. Obowiązkiem administratorów repozytoriów instytucjonalnych powinno być sprawdzenie, czy dane, które mają być opublikowane w otwartym dostępie, składają się z zagregowanych danych badawczych, czy z danych dotyczących indywidualnych uczestników badań[20]. Publikowanie lub udostępnianie tej ostatniej kategorii, tj. danych dotyczących poszczególnych uczestników badań, może powodować ryzyko związane z ponowną identyfikacją. Takie ryzyko może zaistnieć, nawet jeśli są to dane pozbawione elementów pozwalających na identyfikację. Trudno takie ryzyko wykluczyć, szczególnie wobec istnienia zaawansowanych technologii przetwarzania danych.

Środki bezpieczeństwa i rekomendacje dla podmiotów prowadzących repozytoria instytucjonalne

Ważkim problemem jest ustalenie, kto jest administratorem danych osobowych, które wchodzą w skład danych badawczych. Ma to kluczowe znaczenie z uwagi chociażby na obowiązki wynikające z RODO w zakresie informowania uczestników badań naukowych oraz wielu innych obowiązków związanych z ochroną danych osobowych. Z tej perspektywy należy określić rolę badacza, a także podmiotu finansującego badania oraz organizacji realizującej badania, np. szkoły wyższej oraz podmiotu prowadzącego repozytorium instytucjonalne. Unormowania RODO przewidują – oprócz roli administratora danych – także funkcję podmiotu przetwarzającego dane osobowe (tzw. procesora) oraz osoby, które mogą przetwarzać dane osobowe z upoważnienia administratora lub podmiotu przetwarzającego. Ta ostatnia funkcja może w praktyce odpowiadać zadaniom podmiotu prowadzącego repozytorium instytucjonalne, który winien ustalić podmiot będący administratorem danych, ewentualnie ich procesorem

[19] L. Naudts, F. van den Boom, T. Tagarev *et al.*, *Best Practices*…, s. 41.

[20] J. Holzel, *Differential Privacy and the GDPR*, "European Data Protection Law Review" 2019, vol. 5(2), s. 186; por. wyrok TS z 13.05.2014 r. w sprawie C-131/12 *Google Spain and Google*, ECLI:EU:C:2014:317, pkt 53 i n.

oraz uzyskać licencję na ewentualne przechowywanie i udostępnianie danych. W praktyce administratorem danych może być podmiot finansujący badania, sam naukowiec lub ośrodek prowadzący badania, w szczególności uczelnia. Można też wskazać na sytuację, w której wymienione podmioty będą współadministratorami danych lub niektóre z nich będą pełnić rolę podmiotu przetwarzającego dane osobowe. Można także dopuścić sytuację, że sam podmiot finansujący może pełnić jedynie funkcję osoby trzeciej[21].

Unormowania RODO nakładają szereg obowiązków na administratora danych osobowych i procesorów w zakresie bezpieczeństwa danych i ich przetwarzania. Przy tym odpowiedzialność podmiotu prowadzącego repozytorium instytucjonalnego jest oparta na umowie z administratorem danych osobowych[22]. Zakres tej odpowiedzialności zależy zatem od treści samej umowy licencyjnej. Podstawowy ciężar odpowiedzialności w tym względzie spoczywa na administratorze danych osobowych. Wśród najważniejszych obowiązków z zakresu przetwarzania danych osobowych należy wymienić stosowanie reguły minimalizacji przetwarzanych danych. Zgodnie z nią administrator i procesorzy powinni zapewnić, aby dane osobowe były przetwarzane jedynie w niezbędnym zakresie, przez czas wymagany do realizacji celów, dla których te dane są przetwarzane. W analizowanej sytuacji będą to cele badawcze, w tym kwestia weryfikacji poprawności badań. Podstawą do reglamentacji dostępu do danych badawczych obejmujących dane osobowe jest reguła poufności danych, które są przetwarzane. Jej realizacja wiąże się z zabezpieczeniami danych przed dostępem osób nieupoważnionych i udzielaniem tegoż dostępu podmiotom autoryzowanym. Przetwarzane dane osobowe jako element danych badawczych powinny być chronione przed zniszczeniem, nieuprawnionymi zmianami oraz innymi naruszeniami ich integralności. Ten obowiązek jest określany właśnie mianem reguły integralności przetwarzanych danych. Należy zauważyć, że postulat zachowania zupełności i kompletności danych badawczych jest skorelowany z funkcjami repozytorium instytucjonalnego[23].

Podkreślić należy, że administrator ma obowiązek wdrożenia odpowiednich środków technicznych i organizacyjnych, aby przetwarzanie danych odbywało się zgodnie z prawem, przede wszystkim unormowaniami RODO. Adekwatność zastosowanych środków powinna być możliwa do wykazania i podlega kontroli. Zaś zastosowanie niewystarczających czy niewłaściwych środków wiąże się z odpowiedzialnością, co jest określane mianem reguły rozliczalności[24]. Określenie, które ze środków technicznych

---

21 Por. N. Kalinowska, B. Oręziak, M. Świerczyński, *Badania kliniczne w świetle RODO*, „Prawo Mediów Elektronicznych" 2018, nr 3, s. 6; A. Popowicz-Pazdej, *Umowa w przedmiocie współadministrowania danymi osobowymi*, „Prawo Mediów Elektronicznych" 2020, nr 3, s. 20.

22 A. Pyka, *Powierzenie przetwarzania danych osobowych w świetle ogólnego rozporządzenia o ochronie danych*, „Prawo Mediów Elektronicznych" 2020, nr 1, s. 10.

23 K.L. Smouter-Umans, *Research, GDPR, and the DPO How GDPR Changes the Game for Those Conducting Research and the Data Supervisors*, „International Journal for the Data Protection Officer, Privacy Officer and Privacy Counsel" 2017, vol. 1(1), s. 32.

24 P. Litwiński, [w:] *Ogólne rozporządzenie o ochronie danych osobowych. Ustawa o ochronie danych osobowych. Wybrane przepisy sektorowe. Komentarz*, red. *idem*, Legalis, komentarz do art. 5, nb 25; por. wyrok WSA w Warszawie z 2.04.2007 r., II SA/Wa 2328/06, Legalis; A. Mednis, *Ochrona danych osobo-*

i organizacyjnych są odpowiednie, nie jest łatwe. Wydaje się, że powinny być one zdeterminowane takimi okolicznościami, jak cechy danych: poziom wrażliwości i stopień ryzyka istotnej reidentyfikacji. Ale także cel przetwarzania danych. Prowadzenie badań naukowych stanowi, zgodnie z art. 9 RODO, przesłankę dopuszczającą przetwarzanie tzw. wrażliwych danych osobowych. Wskazany przepis określa zatem generalny zakaz przetwarzania danych osobowych wrażliwych. Ta kategoria danych została określona w art. 9 ust. 1 RODO i obejmuje dane osobowe ujawniające pochodzenie rasowe lub etniczne, poglądy polityczne, przekonania religijne lub światopoglądowe, przynależność do związków zawodowych oraz kwestii przetwarzania danych genetycznych i danych biometrycznych, w celu jednoznacznego zidentyfikowania osoby fizycznej, lub danych dotyczących zdrowia, seksualności lub orientacji seksualnej tej osoby. Przetwarzanie danych wrażliwych może być niezbędne np. w celu przeprowadzenia badań klinicznych, socjologicznych, historycznych czy ekonomicznych. Treść art. 89 ust. 1 RODO wskazuje na cele archiwalne w interesie publicznym, związane z realizacją badań naukowych lub historycznych, cele statystyczne jako uzasadnienie przetwarzania danych osobowych z zastrzeżeniem odpowiednich zabezpieczeń. Prawodawca unijny wskazał tu na omówioną wyżej pseudonimizację danych jako jeden ze środków ochrony prawa do prywatności[25].

Należy wskazać, że zgodnie z art. 7 ust. 1–4 RODO, art. 9 ust. 2 lit. a RODO świadoma zgoda uczestników badań na przetwarzanie danych osobowych jest warunkiem legalności tego procesu. W związku z udzieleniem zgody na przetwarzanie danych osobowych aktywuje się szereg obowiązków informacyjnych administratora danych. Świadomej zgodzie uczestnika badań powinny towarzyszyć informacje dotyczące: jego tożsamości i danych kontaktowych uczestnika badań, jak też inspektora ochrony danych, cel przetwarzania danych i jego podstawę prawną, informację o odbiorcach danych osobowych ich kategoriach (np. inni badacze), informacje o zamiarze zdeponowania danych w repozytorium instytucjonalnym lub przekazania do państwa trzeciego czy organizacji międzynarodowej. Jeżeli w procesie przetwarzania danych będą podejmowane decyzje w trybie tzw. zautomatyzowanym, uczestnik badań powinien być poinformowany o zasadach podejmowania takich decyzji, jak również o znaczeniu i przewidywanych konsekwencjach przetwarzania danych[26].

Ponadto, uczestnik badań powinien być poinformowany o okresie, w jakim dane osobowe będą przechowywane i udostępniane w repozytorium instytucjonalnym, lub

---

wych w systemie ochrony zdrowia. Zasady prowadzenia, udostępniania i archiwizowania dokumentacji medycznej [w:] *System Prawa Medycznego*, t. 3, *Organizacja systemu ochrony zdrowia*, Warszawa 2020, s. 941.

[25] M. Kogut-Czarkowska, *Pseudonimizacja i anonimizacja danych osobowych w badaniach naukowych – wybrane zagadnienia*, „Prawo Nowych Technologii" 2021, nr 1, s. 13; por. decyzję Prezesa Urzędu Ochrony Danych Osobowych z 11.06.2019 r., ZSZZS.440.592.2018, Legalis.

[26] E. Marciniak, *Prawo do prywatności a obowiązek zachowania w tajemnicy informacji o pacjencie w kontekście kodeksu etyki lekarskiej oraz RODO*, „Przegląd Prawa Publicznego" 2020, nr 1, s. 62; M. Siwicki, *Ochrona osób fizycznych w związku z przetwarzaniem i swobodnym przepływem danych osobowych (uwagi w związku z projektem rozporządzenia Parlamentu Europejskiego i Rady)*, „Państwo i Prawo" 2016, nr 3, s. 78.

przynajmniej kryteria ustalania tego okresu. Wreszcie, uczestnik badań powinien być poinformowany o swoich prawach, takich jak do żądania dostępu do swoich danych osobowych, ich sprostowania, usunięcia lub ograniczenia przetwarzania, jak również prawie do przenoszenia danych, cofnięcia zgody w dowolnym momencie, prawie wniesienia skargi do organu nadzorczego[27].

Podkreślić przy tym należy, że sama zgoda na przetwarzanie danych osobowych wyrażona przez uczestników badań nie chroni przed ryzykiem ich potencjalnej ponownej identyfikacji. Zgoda ma więc ograniczoną rolę w zapewnieniu poszanowania prywatności i poufności informacji dotyczących osób fizycznych. Natomiast wymóg jej uzyskania stanowi gwarancję legalności przetwarzania danych i realizacji obowiązków ze strony administratora danych osobowych. Pełni też szczególną rolę w zapewnieniu przejrzystości i wiarygodności procesu przetwarzania danych osobowych oraz udostępniania danych badawczych.

A zatem poziom wrażliwości danych, możliwość reidentyfikacji oraz cel przetwarzania danych są czynnikami, które powinny określać poziom dostępu do danych badawczych zdeponowanych w repozytorium instytucjonalnym. Czynniki te powinny mieć – i w praktyce mają – wpływ na szeroki wachlarz mechanizmów bezpieczeństwa danych stosowanych przez repozytoria instytucjonalne. Należy tu wyróżnić typowe poziomy dostępu do danych badawczych. Może to być dostęp otwarty, ograniczony lub kontrolowany.

Środki bezpieczeństwa związane z wrażliwością danych powinny wiązać się z różnymi poziomami kontroli stosowanymi przez administratora danych i realizowanymi także przez repozytorium instytucjonalne. W efekcie zastosowania środków bezpieczeństwa danych dostęp do danych badawczych zawierających dane osobowe powinien być adresowany do działających w dobrej wierze naukowców, którzy byliby zobowiązani postanowieniami umowy o udostępnienie danych badawczych. Jednym z postanowień takiej umowy o udostępnieniu danych może być zobowiązanie podmiotu uzyskującego dostęp do niepodejmowania prób ponownej identyfikacji danych uczestników badań. Innym środkiem bezpieczeństwa jest to, aby dane badawcze były udostępniane za pośrednictwem bezpiecznych platform elektronicznych i nie było możliwe ich pobranie[28].

Pożądany – z perspektywy ochrony prawa do prywatności uczestników badań – jest wymóg stałej współpracy administratorów danych osobowych i depozytariuszy danych badawczych z administratorami repozytorium. Zrozumiałym w tym względzie postulatem jest prowadzenie systematycznego, cyklicznego audytu zabezpieczeń ochrony danych badawczych zawierających dane osobowe.

---

[27] P. Kozik, *Zakres swobody regulacyjnej państw członkowskich przy wdrażaniu ogólnego rozporządzenia o ochronie danych osobowych do prawa krajowego*, „Europejski Przegląd Sądowy" 2017, nr 5, s. 21; M. Wyrwiński, *Nowe obowiązki dostawy treści lub usług cyfrowych. Uwagi wybrane do dyrektywy Parlamentu Europejskiego i Rady (UE) 2019/770*, ZNUJ PPWI 2020, nr 2, s. 249.

[28] D. Dimitrova, *The Right to Explanation under the Right of Access to Personal Data: Legal Foundations in and beyond the GDPR*, "European Data Protection Law Review" 2020, vol. 6(2), s. 214.

Za dobrą praktykę repozytoriów instytucjonalnych należy uznać stosowanie mechanizmu notyfikacji. Może on być ustanowiony poprzez nałożenie na administratorów danych osobowych oraz licencjobiorców obowiązku powiadamiania administratora repozytorium, o wykryciu, że dane osobowe uczestników badań mogą być lub zostały ponownie zidentyfikowane[29]. Po otrzymaniu informacji, że dane badawcze mogą być zagrożone, administrator danych osobowych oraz administrator repozytorium powinni zachować prawo do zawieszenia lub zakończenia udostępniania takich materiałów. Polityka repozytorium instytucjonalnego powinna precyzyjnie zastrzegać prawo do zawieszenia lub zakończenia udostępniania danych[30]. W takim przypadku należy odpowiednio poinformować użytkowników repozytorium instytucjonalnego. W sytuacji naruszenia danych osobowych, administrator danych osobowych, który zdeponował określone dane badawcze i upoważnił podmiot prowadzący repozytorium instytucjonalne do jego udostępniania na określonych warunkach, powinien podjąć wszelkie uzasadnione wysiłki, aby zażądać od wszystkich użytkowników usunięcia całości lub części tych danych, które przykładowo stały się możliwe do ponownego zidentyfikowania. Takie działania powinny obejmować powiadomienia na stronach internetowych, takich jak portale otwartych danych oraz fora internetowe, media społecznościowe czy poczta elektroniczna, do których dostęp mają osoby, które prawdopodobnie ponownie wykorzystują dane. Z jednej strony, wymóg rejestracji użytkowników może być najskuteczniejszym instrumentem zapewnienia skuteczności wycofywania danych badawczych. Z drugiej zaś – może to się wiązać z gromadzeniem nowych danych osobowych od dalszych użytkowników, i w konsekwencji – skutecznie zniechęcać do korzystania z platform elektronicznych prowadzonych w ramach repozytorium instytucjonalnego[31].

Wnioski końcowe

Publiczny dostęp do danych badawczych, które zostały uzyskane przy użyciu środków publicznych jest wartością zasługującą na promowanie. Natomiast decyzja co do publicznego udostępnienia określonych danych badawczych powinna być podejmowana przy uwzględnieniu szeregu okoliczności, takich jak m.in.: ochrona danych osobowych uczestników badań, stopień wrażliwości tych danych, ale także prawo własności intelektualnej i interesy handlowe leżące w możliwej komercjalizacji wyników badań, jak również ochrona innych prawnie chronionych interesów, np. bezpieczeństwa państwa. Ochrona prawa do prywatności osób, których dane są zawarte w wynikach badań

29 L. Naudts, F. van den Boom, T. Tagarev *et al.*, *Best Practices*…, s. 61.

30 A.B. Cordeiro Menezes, *Civil Liability for Processing of Personal Data in the GDPR*, "European Data Protection Law Review" 2019, vol. 5(4), s. 496.

31 N. Gruenpeter, *Jak korzystać z zasobów w repozytoriach danych*, Warszawa 2019, s. 18, https://drodb.icm.edu.pl/wp-content/uploads/2019/10/Jak-korzysta%C4%87-z-zasob%C3%B3w-w-repozytoriach-danych.pdf [dostęp: 9.01.2022].

udostępnianych w repozytoriach instytucjonalnych, stanowi rzeczywisty problem. Niniejszy artykuł stanowi głos w dyskusji na temat założeń Polityki otwartego dostępu do danych badawczych finansowanych ze środków publicznych. Do jej opracowania ustawodawca zobowiązał ministra właściwego do spraw szkolnictwa wyższego i nauki na podstawie art. 23 ust. 1 ustawy z 2021 r. o otwartych danych.

Należy wyrazić opinię, że optymalnym rozwiązaniem jest deponowanie danych badawczych zawierających wrażliwe dane osobowe w repozytorium instytucjonalnym lub jego części o ograniczonym dostępie. Pozwala to administratorom danych osobowych i samym badaczom na ustalenie poziomu dostępu oraz przeprowadzenie odpowiedniej kontroli osób ubiegających się o dostęp do takich danych. Taka kontrola może obejmować sprawdzanie powiązań, kwalifikacji, i poprzez zawarcie umowy o udostępnienie danych, nałożenie zobowiązania, że użytkownik nie będzie udostępniać danych badawczych innym podmiotom, ani próbować reidentyfikować osób fizycznych. Podmioty prowadzące repozytoria instytucjonalne występują i mogą występować w roli podmiotu przetwarzającego dane osobowe (procesora). Mogą to czynić na podstawie i w granicach umowy z administratorem danych osobowych. Z uwagi na daleko idącą odpowiedzialność, podmioty prowadzące repozytoria instytucjonalne nie powinny przejmować roli administratora danych osobowych. Powinna być to domena organizacji prowadzącej badania lub samych badaczy.

Obowiązek udostępniania danych badawczych zgodnie z założeniem „jak to najbardziej możliwe”[32] do zastosowań wtórnych w stosunku do danych badawczych uzyskanych przy wykorzystaniu środków publicznych jest faktem. Dlatego też, tworząc i administrując instytucjonalnym repozytorium danych badawczych, należy zadbać o przejrzystość procesu regulującego dostęp do danych badawczych wrażliwych. Proces ten powinien być transparentny zarówno dla naukowców, administratorów danych osobowych, jak i dla uczestników badań oraz badaczy wnioskujących o dostęp (użytkowników)[33].

Systemy repozytoryjne wykorzystywane do przechowywania i udostępniania danych badawczych powinny nie tylko zapewniać łatwy dostęp do tychże danych, lecz także implementować niezawodne mechanizmy ochrony danych osobowych, w tym danych wrażliwych, przed naruszeniem. Spełnienie tych standardów ochrony leży zarówno w interesie administratorów danych, depozytariuszy danych, jak też podmiotów prowadzących repozytoria instytucjonalne. Przede wszystkim jednak wartością chronioną jest tu prawo do prywatności osób uczestniczących w badaniach. Osoby takie należy uświadomić, że w wyniku ewentualnego otwartego udostępnienia ich danych osobowych wycofanie się z przyszłego wtórnego wykorzystania tych danych dla celów dalszych badań może być trudne lub niemożliwe. Również kwestia zapewnienia otwartego dostępu do danych po procesie pseudonimizacji nie wyklucza całkowicie ryzyka reidentyfikacji.

---

32 Por. przypis 7.

33 Wynika on także z treści art. 7–9 dyrektywy 2019/1024.

Literatura

Arzberger P., Schroeder P., Beaulieu A., *Promoting access to public research data for scientific, economic, and social development*, "Data Science Journal" 2004, nr 3.

Atik J., Jeutner V., *Quantum computing and computational law*, "Law, Innovation and Technology" 2021, vol. 12(2).

Bhuva N., Luca E., *Issues and challenges in researchers' adoption of open access and institutional repositories: a contextual study of a university repository*, "Information Research: an international electronic journal" 2017, vol. 22(4).

Cordeiro Menezes A.B., *Civil Liability for Processing of Personal Data in the GDPR*, "European Data Protection Law Review" 2019, vol. 5(4).

Dawson P.H., Yang S.Q., *Institutional repositories, open access and copyright: what are the practices and implications?*, "Science & Technology Libraries" 2016, vol. 35(4).

Fajgielski P. [w:] *Komentarz do rozporządzenia nr 2016/679 w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych)* [w:] *Ogólne rozporządzenie o ochronie danych. Ustawa o ochronie danych osobowych. Komentarz*, LEX/el.

Fischer B., *Autorskoprawne konteksty ponownego wykorzystania danych badawczych* [w:] *Sto lat polskiego prawa handlowego. Księga jubileuszowa dedykowana Profesorowi Andrzejowi Kidybie*, red. M. Dumkiewicz, K. Kopaczyńska-Pieczniak, J. Szczotka, t. 2, Warszawa 2020.

Hintze M., El Emam K., *Comparing the benefits of pseudonymisation and anonymisation under the GDPR*, "Journal of Data Protection & Privacy" 2018, vol. 2(2).

Holzel J., *Differential Privacy and the GDPR*, "European Data Protection Law Review" 2019, vol. 5(2).

Kalinowska N., Oręziak B., Świerczyński M., *Badania kliniczne w świetle RODO*, „Prawo Mediów Elektronicznych" 2018, nr 3.

Kogut-Czarkowska M., *Pseudonimizacja i anonimizacja danych osobowych w badaniach naukowych – wybrane zagadnienia*, „Prawo Nowych Technologii" 2021, nr 1.

Kozik P., *Zakres swobody regulacyjnej państw członkowskich przy wdrażaniu ogólnego rozporządzenia o ochronie danych osobowych do prawa krajowego*, „Europejski Przegląd Sądowy" 2017, nr 5.

Kuner C., Cate F., Lynskey O., Millard C., Loideain N.N., Svantesson D., *Blockchain versus data protection*, "International Data Privacy Law" 2018, vol.8(2).

Litwiński P. [w:] *Ogólne rozporządzenie o ochronie danych osobowych. Ustawa o ochronie danych osobowych. Wybrane przepisy sektorowe. Komentarz*, red. *idem*, Legalis.

Lubasz D., Chomiczewski W., Czerniawski, M., Drobek, P., Góral, U., Kuba M., Makowski, P., Witkowska-Nowakowska K. [w:] *RODO. Ogólne rozporządzenie o ochronie danych. Komentarz*, red. E. Bielak-Jomaa, LEX/el.

Marciniak E., *Prawo do prywatności a obowiązek zachowania w tajemnicy informacji o pacjencie w kontekście kodeksu etyki lekarskiej oraz RODO*, „Przegląd Prawa Publicznego" 2020, nr 1.

Markiewicz R., *Prawo autorskie na jednolitym rynku cyfrowym. Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, Warszawa 2021.

Mednis A., *Ochrona danych osobowych w systemie ochrony zdrowia. Zasady prowadzenia, udostępniania i archiwizowania dokumentacji medycznej* [w:] *Organizacja systemu ochrony zdrowia. System Prawa Medycznego*, t. 3, Warszawa 2020.

Miller W., *Nowa ustawa o ponownym wykorzystywaniu informacji sektora publicznego*, Temidium 2021, nr 3.

Ossowska-Salamonowicz D., *Ochrona danych osobowych w działalności dziennikarskiej*, Olsztyn 2015.

Popowicz-Pazdej A., *Umowa w przedmiocie współadministrowania danymi osobowymi*, „Prawo Mediów Elektronicznych" 2020, nr 3.

Pyka A., *Powierzenie przetwarzania danych osobowych w świetle ogólnego rozporządzenia o ochronie danych*, „Prawo Mediów Elektronicznych" 2020, nr 1.

Siwicki M., *Ochrona osób fizycznych w związku z przetwarzaniem i swobodnym przepływem danych osobowych (uwagi w związku z projektem rozporządzenia Parlamentu Europejskiego i Rady)*, „Państwo i Prawo" 2016, nr 3.

Smouter-Umans K.L., *Research, GDPR, and the DPO How GDPR Changes the Game for Those Conducting Research and the Data Supervisors*, "International Journal for the Data Protection Officer, Privacy Officer and Privacy Counsel" 2017, vol. 1(1).

Więckowska M., *Anonimizacja i pseudonimizacja to główne narzędzia pracy administratorów danych, podwyższające bezpieczeństwo danych oraz osób, których dane są udostępniane i przetwarzane*, „ABI Expert" 2016, nr 1.

Wyrwiński M., *Nowe obowiązki dostawy treści lub usług cyfrowych. Uwagi wybrane do dyrektywy Parlamentu Europejskiego i Rady (UE) 2019/770*, ZNUJ. PPWI 2020, nr 2.

Streszczenie

Marek Salamonowicz

Otwarty dostęp do danych badawczych w ramach repozytorium instytucjonalnego a ochrona danych osobowych i prawo do prywatności

Artykuł stanowi głos w dyskusji co do zasad Polityki otwartego dostępu do danych badawczych finansowanych ze środków publicznych, która z mocy ustawy z 11 sierpnia 2021 r. o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego ma być opracowana przez ministra właściwego ds. szkolnictwa wyższego i nauki i przedstawiona jednostkom naukowym. W materiale sformułowano szereg rekomendacji w zakresie otwartego dostępu do danych badawczych. Podmioty prowadzące repozytoria instytucjonalne danych badawczych, w szczególności uczelnie, powinny podejmować we współpracy z administratorami danych osobowych szereg środków organizacyjnych i technicznych w celu zapewnienia poszanowania prawa do prywatności uczestników badań naukowych. Oprócz anonimizacji i pseudonimizacji danych, w uzasadnionych przypadkach należy ograniczać dostęp do danych badawczych obejmujących dane osobowe. Czynnikami, które powinny wpływać na zastosowany poziom ograniczeń są m.in.: wrażliwość danych, ryzyko ich reidentyfikacji. Zasady takiego kontrolowanego dostępu powinny być transparentne zarówno dla badaczy, potencjalnych użytkowników, jak też administratorów danych osobowych i samych uczestników badań naukowych. Optymalizacja procesu dostępu do danych badawczych powinna pozytywnie wpłynąć na rozwój badań, przy zapewnieniu należytej ochrony prawa do prywatności.

Słowa kluczowe: dane badawcze; dane osobowe; otwarty dostęp; pseudonimizacja; anonimizacja; reidentyfikacja.

Summary

Marek Salamonowicz

Open Access to Research Data in an Institutional Repository and Protection of Personal Data and Right to Privacy

The article is a voice in a discussion about the principles of the Open Access Policy for publicly funded research data, which under the Act of 11 August 2021 on open data and reuse of public sector information is to be developed by the minister responsible for higher education and science and presented to scientific entities. The material makes several recommendations for open access to research data. Entities maintaining institutional repositories of research data, in particular universities, should take a number of organisational and technical measures in cooperation with personal data controllers to ensure respect for the right to privacy of research participants. In addition to data anonymisation and pseudonymisation, access to research data involving personal data should be restricted where justified. Factors that should influence the level of restrictions include the sensitivity of the data and the risk of re-identification. The principles of such controlled access should be transparent for researchers, potential users, as well as personal data administrators, and research participants themselves. The optimisation of the process of access to research data should have a positive impact on the development of research while ensuring the adequate protection of the right to privacy.

Keywords: research data; personal data; open access; pseudonymisation; anonymisation; re-identification.