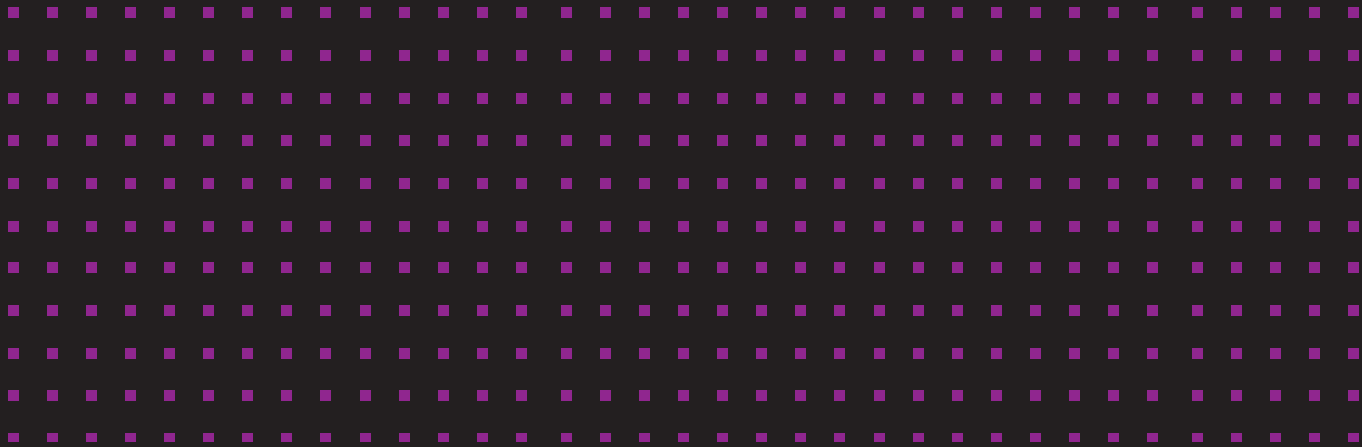


# Otwarty dostęp w instytucjach naukowych



# Otwarty dostęp w instytucjach naukowych

pod redakcją

Jakuba Szprota

autorzy

Wojciech Fenrich

Tomasz Lewandowski

Krzysztof Siewicz

Michał Starczewski

Lidia Stępińska-Ustasiak

Jakub Szprot

Wydawnictwa ICM

Warszawa 2015

Redakcja i korekty  
Maciej Chojnowski

Projekt okładki, stron tytułowych, opracowanie graficzne, skład i łamanie  
Jakub Rakusa-Suszczewski

© Copyright by Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego, Warszawa 2014. Pewne prawa zastrzeżone. Tekst niniejszej publikacji jest dostępny na licencji Creative Commons – Uznanie autorstwa 3.0 Polska. Postanowienia licencji są dostępne pod <http://creativecommons.org/licenses/by/3.0/pl/legalcode>.

ISBN 978-83-63490-11-9

Wydawca

Wydawnictwa ICM  
ul. Pawińskiego 5a  
02-106 Warszawa



Niniejszy raport został przygotowany w ramach działań Platformy Otwartej Nauki



## Spis treści

Wstęp .....	6
<b>CZĘŚĆ I</b>	
Rozwiązania światowe i lokalne okoliczności .....	9
<b>ROZDZIAŁ PIERWSZY</b>	
Polityki otwartego dostępu w świecie .....	10
1.1. Polityki otwartego dostępu i otwarte mandaty .....	10
1.2. Uniwersytet Harvarda .....	12
1.3. Uniwersytet w Liège .....	14
1.4. Węgierska Akademia Nauk .....	15
1.5. Politechnika Federalna w Zurichu .....	16
1.6. Podsumowanie .....	17
<b>ROZDZIAŁ DRUGI</b>	
Kontekst polski – przekonania, praktyki, oczekiwania .....	19
2.1. Forma i rola publikacji naukowych .....	19
2.2. Forma i funkcje otwartego dostępu .....	22
2.3. Najważniejsze wnioski w kontekście kształtowania i wdrażania polityki otwartego dostępu .....	27
<b>CZĘŚĆ II</b>	
Otwarty dostęp w polskich instytucjach naukowych – rekomendacje .....	28
<b>ROZDZIAŁ TRZECI</b>	
Polityka otwartości .....	29
3.1. Rodzaje otwartego dostępu .....	30
3.2. Polityka otwartości a prawo autorskie .....	30

3.2.1. Autorskie prawa osobiste .....	31
3.2.2. Autorskie prawa majątkowe .....	32
3.3. Polityka otwartości a praktyki wydawnicze .....	32
3.4. Polityka otwartości w hierarchii wewnętrznych aktów prawnych na uczelni .....	34

## ROZDZIAŁ CZWARTY

Działalność wydawnicza .....	37
4.1. Cele działalności wydawniczej .....	37
4.1.1. Wspieranie pracowników naukowych w budowaniu pozycji, prestiżu i spełnianiu wymagań instytucji finansujących badania .....	38
4.1.2. Wspieranie instytucji naukowej we wzmacnianiu jej pozycji w środowisku akademickim i poza nim .....	38
4.1.3. Hierarchizacja celów i korzyści .....	39
4.2. Kwestie prawnoautorskie .....	39
4.2.1. Stosunki prawnoautorskie między wydawcą a autorami .....	40
4.3. Kwestie finansowe .....	41
4.3.1. Opłaty za publikacje .....	41
4.4. Kwestie techniczne związane z udostępnianiem publikacji elektronicznych .....	41
4.4.1. Narzędzia informatyczne dostosowane do typu publikacji .....	42
4.4.1.1. Open Journal Systems .....	42
4.4.1.2. Open Monograph Press .....	42
4.4.1.3. Bazy czasopism .....	43
4.4.2. Metadane .....	43
4.4.2.1. DOI (Digital Object Identifier) .....	43
4.4.3. Interoperacyjność .....	44
4.4.4. Widoczność w Internecie .....	44
4.4.5. Enhanced publications .....	44
4.5. Czasopisma a monografie .....	45
4.6. Publikacje bieżące a publikacje archiwalne .....	46
4.6.1. Prawne aspekty udostępniania starszych publikacji .....	46
4.6.2. Postać publikacji .....	46
4.7. Podsumowanie .....	47

## ROZDZIAŁ PIĄTY

Repozytoria .....	48
5.1. Repozytoria jako część systemu udostępniania publikacji naukowych w Internecie .....	49
5.2. Europejska infrastruktura otwartego dostępu OpenAIRE .....	51
5.3. Wyszukiwarki a treści naukowe w Internecie .....	52
5.3.1. Ogólna charakterystyka wyszukiwarek internetowych .....	52
5.3.2. Specyfika repozytoriów jako obiektów indeksowanych przez wyszukiwarki internetowe .....	53

5.4. Metadane .....	55
5.4.1. HTML .....	55
5.4.2. Dane bibliograficzne w rekordzie repozytorium .....	57
5.4.3. Ekstrakcja metadanych .....	60
5.5. Rekomendacje dla repozytoriów .....	61
5.5.1. Wybór oprogramowania .....	61
5.5.2. Widoczność zasobów repozytorium w wyszukiwarkach internetowych .....	62
5.5.2.1. Mapa strony .....	62
5.5.2.2. Pliki robots.txt .....	63
5.5.2.3. Przeglądanie .....	63
5.5.2.4. Brak stron pośrednich między rekordem a pełnym tekstem publikacji .....	63
5.5.2.5. Linki .....	64
5.5.2.5.1. Linki w mediach społecznościowych .....	64
5.5.2.5.2. Linki w Wikipedii .....	65
5.5.2.6. Pełne teksty artykułów dostępne lokalnie .....	65
5.5.3. Metadane .....	66
5.5.3.1. Format metadanych .....	66
5.5.3.2. Metadane dotyczące autorów a nazwiska widoczne w artykule .....	67
5.5.3.3. Metadane dotyczące daty .....	69
5.5.3.4. Cytowania .....	69
5.5.4. Wybrane elementy architektury strony .....	73
5.5.4.1. Optymalizacja wydajności strony .....	73
5.5.4.2. Optymalizacja bazy danych repozytoriów .....	74
5.5.4.3. Przyjazne użytkownikowi adresy URL .....	74
5.5.4.4. Komunikaty o błędach .....	74
5.5.4.4.1. Własne komunikaty o błędach .....	75
5.5.4.4.2. Miękkie 404 .....	75
5.5.4.4.3. Inne komunikaty o błędach.....	75

## ROZDZIAŁ SZÓSTY

Działania informacyjne i komunikacyjne .....	76
6.1. Wprowadzenie otwartego dostępu w instytucji .....	77
6.1.1. Polityka otwartości .....	77
6.1.2. Repozytorium instytucjonalne .....	78
6.2. Polityka otwartości w działaniu .....	78
6.3. Podsumowanie .....	80
Bibliografia .....	81
Spis ilustracji .....	85
Spis tabel .....	85

## Wstęp

W podsumowaniu opublikowanego w ubiegłym roku raportu *Otwarta nauka w Polsce 2014. Diagnoza* podkreślaliśmy, że otwarty dostęp do treści naukowych jest w Polsce realizowany przede wszystkim przez wydawców czasopism naukowych, samych badaczy, ośrodki oferujące rozwiązania informatyczne oraz nieliczne instytucje prowadzące repozytoria. Brakuje natomiast instytucjonalnych strategii i polityk otwartego dostępu – na poziomie rządowym, instytucji finansujących badania oraz instytucji naukowych<sup>1</sup>. W ostatnim czasie pewne działania w tym zakresie podjęło Ministerstwo Nauki i Szkolnictwa Wyższego, które 20 marca 2015 roku powołało Zespół doradczy do spraw otwartego dostępu do treści naukowych<sup>2</sup>. Do zadań Zespołu należało przygotowanie *Planu wdrożenia otwartego dostępu do treści naukowych w Polsce* oraz analiza nadesłanych uwag i komentarzy, a także opracowanie projektu *Kierunków rozwoju otwartego dostępu do treści naukowych w Polsce* i przedłożenie tego projektu do akceptacji Ministrowi Nauki i Szkolnictwa Wyższego. W październiku 2015 roku Minister przyjął dokument *Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce*<sup>3</sup>. Zawiera on między innymi zalecenia dla instytucji naukowych, którym rekomenduje się przyjęcie i wdrożenie polityki otwartego dostępu:

Ministerstwo Nauki i Szkolnictwa Wyższego zaleca jednostkom naukowym i uczelniom zapewnienie otwartego dostępu (z preferencją dla dostępu libre) do publikacji oraz dostępu do danych badawczych powstających dzięki finansowaniu lub współfinansowaniu ze środków publicznych. W tym celu konieczne jest opracowanie i przyjęcie instytucjonalnych polityk w zakresie otwartego dostępu na szczeblu jednostek naukowych lub uczelni. Polityki te w szczególności mogą zawierać zobowiązanie twórców publikacji naukowych powstających w wyniku finansowania ze środków publicznych do zamieszczenia ich w konkretnym repozytorium (...). Zobowiązanie to może być realizowane w oparciu o regulamin, o którym mowa w art. 86c ust. 1 pkt 1 ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (t.j. Dz. U. z 2012 r. poz. 572, z późn. zm.)<sup>4</sup>.

1 *Otwarta nauka w Polsce 2014. Diagnoza*, red. J. Szprot, Wydawnictwa ICM, Warszawa 2014, s. 103.

2 Zob. Zarządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 20 marca 2015 r. w sprawie powołania Zespołu doradczego do spraw otwartego dostępu do treści naukowych, [http://www.nauka.gov.pl/g2/oryginal/2015\\_05/f0061d2ae21e462a5816f8a8cbe4fdb.pdf](http://www.nauka.gov.pl/g2/oryginal/2015_05/f0061d2ae21e462a5816f8a8cbe4fdb.pdf).

3 *Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce*, [http://www.nauka.gov.pl/g2/oryginal/2015\\_10/9f62cc350837b942e51ae23dd1f23df8.pdf](http://www.nauka.gov.pl/g2/oryginal/2015_10/9f62cc350837b942e51ae23dd1f23df8.pdf).

4 Tamże, s. 12.

MNiSW zwraca uwagę na potrzebę tworzenia repozytoriów instytucjonalnych oraz konieczność prowadzenia przez instytucje naukowe działań informacyjnych, promocyjnych i szkoleniowych w zakresie otwartego dostępu, skierowanych do pracowników naukowych. Wydawcom czasopism naukowych uzyskujących finansowanie ze środków publicznych również zalecane jest stosowanie otwartego dostępu w modelu libre.

Polityki otwartego dostępu są przyjmowane na całym świecie. Serwis The Registry of Open Access Repository Mandates and Policies (<http://roarmap.eprints.org>) odnotowuje 724 instytucjonalne polityki dotyczące otwartego dostępu<sup>5</sup>. Spośród nich 79 to polityki instytucji finansujących badania naukowe, 54 – instytucji zarówno finansujących, jak i prowadzących badania, 512 – instytucji naukowych (np. uniwersytetów czy instytutów badawczych), 71 – jednostek wchodzących w skład instytucji naukowych, 8 – organizacji zrzeszających instytucje naukowe.

Głównym powodem tak powszechnego przyjmowania i wdrażania polityk otwartości jest rozpoznanie korzyści, z jakimi się one wiążą – tak dla instytucji naukowych, jak i ich pracowników. Otwarty dostęp zwiększa widoczność i łatwość wyszukiwania publikacji składających się na dorobek naukowy danej instytucji, a co za tym idzie – prowadzi do zwiększenia liczby cytowań i oddziaływania tych publikacji; ułatwia i przyspiesza komunikację naukową oraz współpracę między różnymi podmiotami – zarówno z obszaru nauki, jak i spoza niego (np. naukowców z przedsiębiorcami); pomaga w zarządzaniu wiedzą i przeprowadzaniu procedur ewaluacyjnych; umożliwia racjonalizację kosztów dystrybucji wiedzy; ułatwia archiwizację i długoterminowe przechowywanie treści w postaci cyfrowej; jest wreszcie skutecznym narzędziem promocji.

Na gruncie polskim wyrazem świadomości zalet otwartych modeli dystrybucji wiedzy wśród instytucji naukowych jest Stanowisko Prezydium KRASP i Prezydium PAN z dnia 5 lipca 2013 r. w sprawie zasad otwartego dostępu do treści publikacji naukowych i edukacyjnych. Oba gremia uznają zasadę otwartego dostępu za „sprzyjającą poszerzaniu zakresu współpracy naukowej i skali wykorzystywania wyników badań, w szczególności prowadzących do powstawania innowacyjnych rozwiązań, a także zapewniającą współdziałanie społeczeństwa w formułowaniu nowych wyzwań badawczych”, opowiadają się za tzw. zieloną drogą otwartego dostępu oraz wyrażają wolę współpracy na rzecz wdrażania tego modelu<sup>6</sup>. Jednoznaczne stanowisko KRASP i PAN jest tym istotniejsze, że dotychczas tylko dwie instytucje naukowe w Polsce – Politechnika Śląska oraz Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego UW – przyjęły silną politykę otwartego dostępu w formie tzw. otwartych mandatów, czyli prawnego zobowiązania pracowników do publikowania prac naukowych w otwartym dostępie. Otwartość wciąż nie jest elementem strategii większości instytucji naukowych, co grozi zaprzepaszczaniem potencjału oddolnych działań i inicjatyw.

---

<sup>5</sup> Stan z 7 września 2015 r.

<sup>6</sup> Stanowisko Prezydium KRASP i Prezydium PAN z dnia 5 lipca 2013 r. w sprawie zasad otwartego dostępu do treści publikacji naukowych i edukacyjnych, [http://www.aktualnosci.pan.pl/images/stories/pliki/2013/Wiadomości\\_biez/stanowisko\\_KRASP-PAN\\_open\\_access.pdf](http://www.aktualnosci.pan.pl/images/stories/pliki/2013/Wiadomości_biez/stanowisko_KRASP-PAN_open_access.pdf).



Korzyści, jakie płyną z otwartego dostępu, nie są jedynym argumentem na rzecz uwzględnienia tego modelu przez polskie instytucje naukowe. Instytucje te muszą również brać pod uwagę przemiany zachodzące w ich europejskim i światowym otoczeniu oraz polityki instytucji finansujących badania, przede wszystkim Komisji Europejskiej. Polscy uczeni i zespoły badawcze mają w swojej pracy coraz częściej do czynienia z nowymi formami upowszechniania rezultatów badań; do zadań ich macierzystych instytucji powinno należeć zapewnienie im narzędzi i kompetencji niezbędnych do sprawnego funkcjonowania w dynamicznie zmieniającym się systemie dystrybucji wiedzy.

W niniejszym raporcie omawiamy podstawowe zagadnienia związane z wdrażaniem otwartego dostępu w instytucjach naukowych. Część pierwsza zarysowuje światowe i lokalne tło: w rozdziale pierwszym przedstawiamy studia przypadków dotyczące czterech zagranicznych instytucji naukowych, które przyjęły politykę otwartości, a w rozdziale drugim omawiamy wyniki badania przeprowadzonego metodą zogniskowanego wywiadu grupowego z udziałem osób, których praca związana jest z tworzeniem oraz upowszechnianiem wiedzy naukowej. W części drugiej przedstawiamy rekomendacje dotyczące najważniejszych aspektów wprowadzenia otwartego dostępu na poziomie instytucjonalnym: polityki otwartości i otwartego mandatu (rozdział trzeci), działalności wydawniczej (rozdział czwarty), repozytoriów instytucjonalnych (rozdział piąty) oraz działań informacyjnych i komunikacyjnych (rozdział szósty).

# **I. Rozwiązania światowe i lokalne okoliczności**

# Rozdział 1

## Polityki otwartego dostępu na świecie

### 1. 1. Polityki otwartego dostępu i otwarte mandaty

Polityki instytucji naukowych w odniesieniu do kwestii otwartego dostępu do publikacji naukowych w ogólnym ujęciu podzielić można na *rezolucje*, w których instytucja jedynie zaleca udostępnianie w modelu otwartym publikacji stanowiących rezultat prowadzonych w niej badań naukowych, nierzadko nie precyzując przy tym swych oczekiwań odnośnie do momentu czy sposobu udostępniania treści, oraz bardziej jednoznaczne i stanowcze *otwarte mandaty*, w których instytucje finansujące badania lub instytucje naukowe zobowiązują określone osoby do działań mających zapewnić otwarty dostęp do wybranych rodzajów publikacji naukowych – przede wszystkim recenzowanych artykułów prezentujących oryginalne wyniki badań naukowych<sup>7</sup>. Wyrażenie „otwarty mandat” jest przy tym kalką z języka angielskiego, gdzie „open mandate” oznacza właśnie zobowiązanie do udostępniania publikacji naukowych w sposób otwarty, nakładane na pracowników i/lub (rzadziej) studentów danej instytucji. „Otwarty mandat” znaczy więc tyle, co „obowiązek otwartości”.

Mówiąc o otwartym mandacie, bardzo łatwo przeoczyć fakt, iż celem dokumentów tego rodzaju jest przede wszystkim ułatwienie naukowcom negocjacji z innymi podmiotami zaangażowanymi w proces publikowania i dystrybucji wyników badań – przede wszystkim z wydawcami – które mogłyby postrzegać otwarty dostęp jako niekorzystny dla nich samych, a których pozycja negocjacyjna w sytuacji braku takiego zobowiązania jest znacznie silniejsza od pozycji autora lub autorów. Innymi słowy, otwarte mandaty nie tylko dyscyplinują samych naukowców, wprost zobowiązując ich do udostępniania publikacji w modelu otwartym, ale też wzmacniają ich pozycję negocjacyjną, umożliwiając im uzyskanie pożądanego sposobu dystrybucji treści, który jest korzystny również dla nich samych.

O dostrzeganiu przez samych autorów korzyści płynących z otwartego dostępu świadczy m.in. fakt, że w Stanach Zjednoczonych znaczna część otwartych mandatów zaczyna obowiązywać w rezultacie jednomyślnych głosowań przeprowadzanych wśród kadry naukowej lub ciał złożonych z jej przedstawicieli

---

<sup>7</sup> P. Suber, *Otwarty dostęp*, przeł. R. Bogacewicz et. al., WUW, Warszawa 2014, s. 70. Dostępny w Internecie pod adresem: <http://pon.edu.pl/index.php/nasze-publicacje?pubid=14>.

(takich jak uczelniane senaty lub ich miejscowe odpowiedniki)<sup>8</sup>, a nie w wyniku odgórnych decyzji władz.

Negocjując postanowienia umowy wydawniczej, autor objęty otwartym mandatem może powołać się na ciążące na nim zobowiązanie do zapewnienia otwartego dostępu do publikacji i zażądać odpowiedniego skorygowania kształtu umowy, np. za pomocą załącznika autorskiego. Jest to towarzyszący często otwartemu mandatowi wzorcowy dokument, który po akceptacji przez wydawcę i autora zmienia postanowienia wiążącej ich umowy w sposób pozwalający autorowi na zapewnienie otwartego dostępu do publikacji na zasadach określonych w obejmującym go zobowiązaniu. Wydawca może w tej sytuacji zaakceptować modyfikacje wynikające z postanowień mandatu, obstawać przy swojej wersji umowy lub odmówić publikacji tekstu. Wybierając tę ostatnią możliwość, rezygnuje jednak tym samym z wartościowych treści, co w dłuższej perspektywie nie pozostanie bez wpływu na prestiż i poczytność wydawanych przez niego czasopism. Jeśli wydawca wybiera drugie rozwiązanie, większość otwartych mandatów daje autorom możliwość ubiegania się o zwolnienie z całości lub części zobowiązań. Możliwość ta stanowi jednak ukłon nie tyle w stronę niechętnych otwartemu dostępowi wydawców, co w stronę chcących zachować pełną swobodę wyboru miejsca publikacji autorów. Świadczy o tym np. fakt, iż odsetek próśb o tzw. *waiver*, czyli zdjęcie z konkretnej publikacji obowiązku udostępnienia jej w sposób otwarty, w Massachusetts Institute of Technology wynosi mniej niż 2%, a w przypadku Uniwersytetu Harvarda – poniżej 5%<sup>9</sup>.

Peter Suber wyróżnia następujące ogólne modele otwartych mandatów<sup>10</sup>:

- Mandaty z furtką – wymagają od autorów otwarcia artykułów za pośrednictwem zielonej drogi, o ile pozwala na to ich wydawca. Jeśli takiej zgody brak, autor nie musi deponować tekstu w repozytorium.
- Mandaty depozytowe – wymagają od autora archiwizacji tekstu w momencie jego przyjęcia do publikacji. Tekst taki staje się dostępny w modelu otwartym wtedy, kiedy pozwala na to jego wydawca. Jeśli takiej zgody brak, tekst pozostaje zamknięty, co nie zwalnia jednak autora z obowiązku zdeponowania jego kopii.
- Mandaty oparte na zachowaniu praw – bazują na licencji z góry udzielonej przez autora jego instytucji macierzystej, która zyskuje w ten sposób prawo do przechowywania i udostępniania tekstów w modelu otwartym. W praktyce znaczy to, iż dostępność tekstu w sposób otwarty jest sytuacją domyślną, która nie wymaga uzyskiwania dodatkowych zgód i która może ulec zmianie jedynie na mocy wystosowanej przez autora prośby o uchylenie obowiązku otwartości przez instytucję realizującą otwarty mandat. Ze względu na różnice w prawie autorskim, rozwiązanie to – szczególnie popularne w USA – może być jednak niemożliwe do zastosowania w niektórych kontekstach prawnych.

Jak zauważa Suber, powyższe rodzaje otwartych mandatów mają charakter typów idealnych, od których rzeczywiste przypadki – w tym opisane poniżej – w mniejszym lub większym stopniu odbiegają. Rzeczywiste mandaty mogą różnić się między sobą zarówno w odniesieniu do mechanizmów zapewnienia otwartego

8 Pełną listę takich jednogłośnie głosowań znaleźć można na stronie Open Access Directory: [http://oad.simmons.edu/oadwiki/Unanimous\\_faculty\\_votes](http://oad.simmons.edu/oadwiki/Unanimous_faculty_votes).

9 P. Suber, *Otwarty dostęp*, op. cit., s. 73.

10 Tamże, s. 70–71.

dostępu do publikacji, jak i stopnia szczegółowości rozstrzygnięć dotyczących np. tego, w którym momencie winna nastąpić archiwizacja publikacji, jaka wersja publikacji ma zostać zarchiwizowana czy w jakim serwisie należy takiej archiwizacji dokonać.

Mandaty różnią się również odnośnie do tego, kto – jaka instytucja lub jaka jednostka danej instytucji – i w jaki sposób winien od strony technicznej czy prawnej wspierać osoby objęte mandatem. Regułą jest jednak to, że takie wsparcie istnieje, co na instytucję otwartego mandatu każe spojrzeć nie tylko przez pryzmat koniecznych działań technicznych (np. uruchomienie i prowadzenie repozytorium), prawnych (opracowanie i udostępnienie załączników autorskich ułatwiających negocjacje z wydawcami) czy finansowych, ale również – jeśli nie przede wszystkim – szeregu „miękkich” działań edukacyjnych, promocyjnych i wspierających. Głównymi beneficjentami tych działań winny być każdorazowo te osoby afiliowane w instytucji wdrażającej mandat, których bezpośrednio on dotyczy.

Poniżej omówione zostaną cztery przykłady polityk otwartości przyjętych przez zagraniczne instytucje naukowe. Zostały one dobrane tak, by prezentowały rozwiązania zróżnicowane zarówno pod względem geograficznym (USA, Europa Zachodnia, Europa Środkowa), jak i ze względu na rodzaj instytucji wprowadzającej obowiązek otwartości (uniwersytet, uczelnia techniczna, akademia nauk), a także charakter rozwiązań wykorzystywanych w celu zapewnienia otwartego dostępu do publikacji stanowiących dorobek poszczególnych instytucji. Mogą one stanowić inspirację dla rozwiązań wdrażanych na gruncie polskim. Rekomendacjom dotyczącym otwartego mandatu w kontekście polskiego prawa poświęcony jest rozdział trzeci niniejszego raportu.

## 1.2. Uniwersytet Harvarda

Uniwersytet Harvarda realizuje politykę otwartości od 2008 roku, kiedy to pierwszy z jego wydziałów – Harvard Faculty of Arts and Sciences – przyjął silną politykę otwartego dostępu, która stała się rozwiązaniem modelowym tak dla pozostałych jednostek tej uczelni, jak i dla kilkudziesięciu innych amerykańskich instytucji naukowych, m.in. MIT czy Stanford<sup>11</sup>. Według stanu na koniec 2015 roku wszystkie dziewięć wydziałów oraz trzy harwardzkie centra badawcze (Berkman Center for Internet and Society, Shorenstein Center on Media, Politics and Public Policy oraz Harvard University Center for the Environment) posiadały politykę otwartego dostępu.

Swój pionierski charakter polityka otwartego dostępu Harvard Faculty of Arts and Sciences zawdzięcza przede wszystkim sposobowi, w jaki została przyjęta. Jako pierwsza polityka otwartego dostępu weszła ona w życie w wyniku (w tym przypadku – jednomyślnego) głosowania samych członków społeczności naukowej. Pozwoliło to uniknąć wrażenia, jakoby otwarty mandat stanowił zobowiązanie nakładane z góry na niechętnych otwartemu dostępowi naukowców, a nie narzędzie wspierające model dystrybucji treści naukowych korzystny również dla samych badaczy. Na wynik głosowania duży wpływ miała kampania informacyjna i promocyjna,

---

11 Pełną listę uczelni wzorujących się w tym zakresie na Harvardzie znaleźć można na stronie internetowej Harvard Open Access Project: [http://cyber.law.harvard.edu/hoap/Additional\\_resources#Policies\\_of\\_the\\_kind\\_recommended\\_in\\_the\\_guide](http://cyber.law.harvard.edu/hoap/Additional_resources#Policies_of_the_kind_recommended_in_the_guide).

w której główną rolę odegrał Stuart Shieber, harwardzki profesor informatyki i główny orędownik mandatu. Zgodnie z przyjętą przez poszczególne wydziały Harvardu polityką otwartego dostępu, każdy pracownik naukowy przyznaje uczelni niewyłączne, nieodwołalne prawo do dystrybucji swoich artykułów naukowych w dowolnym niekomercyjnym celu. By tę dystrybucję umożliwić, pracownicy naukowcy winni dostarczyć elektroniczną kopię ostatecznej wersji autorskiej artykułu nie później niż w momencie jego publikacji. W przypadku późniejszego przeniesienia przez pracownika majątkowych praw autorskich na inny podmiot – np. na któregoś z wydawców (a znaczna część z nich tego właśnie wymaga) – zgodnie z amerykańskim prawem autorskim licencja, o ile zostanie udzielona na piśmie, będzie mieć pierwszeństwo przed umową przeniesienia praw. Dlatego Uniwersytet prosi swoich pracowników o to, by dodatkowo potwierdzili oni treść otwartego mandatu podpisując własnoręcznie papierową wersję licencji<sup>12</sup>.

Każdy członek kadry naukowej Uniwersytetu Harvarda objęty otwartym mandatem uczelni może zwrócić się do jej władz z prośbą o zwolnienie z obowiązku zapewnienia otwartego dostępu do określonej publikacji. Prośba taka musi być złożona na piśmie, umotywowana i dotyczyć konkretnej publikacji (nie jest możliwe zwolnienie wszystkich publikacji danego autora *a priori*). Ponadto może ona dotyczyć jedynie zwolnienia z obowiązku udostępnienia publikacji w sposób otwarty, ale nie z samego obowiązku depozytu. Dzięki temu w sytuacji, gdy w przyszłości udostępnienie danej publikacji stanie się możliwe, uczelnia będzie już dysponować jej odpowiednią cyfrową kopią, dzięki czemu nie będą konieczne żadne specjalne zabiegi natury technicznej. Co więcej, uzyskane zwolnienie dotyczy jedynie obowiązku otwarcia publikacji na zasadach przewidzianych w polityce uczelni i nie wyklucza zapewnienia otwartego dostępu na innych warunkach, np. po upływie okresu embargo przewidzianego przez wydawcę. Niezależnie od dostępności pełnego tekstu artykułu jego metadane są przez cały czas dostępne w repozytorium, co pozwala na ich szybką indeksację i upowszechnianie informacji o publikacji.

Wspólnym dla wszystkich wydziałów serwisem służącym udostępnianiu dorobku pracowników Uniwersytetu Harvarda jest repozytorium instytucjonalne DASH, prowadzone przez wyodrębnioną jednostkę w obrębie biblioteki uniwersyteckiej – Office for Scholarly Communication. Repozytorium to jest ponadto jednym z miejsc przechowywania i upowszechniania prac licencjackich i magisterskich oraz rozpraw doktorskich studentów i doktorantów Uniwersytetu Harvarda. Według stanu z grudnia 2015 roku, serwis od swego uruchomienia w roku 2009 zanotował ponad 6,6 miliona pobrań dokumentów, których w bazie znajduje się już ponad 27 tysięcy<sup>13</sup>.

Badacze z Uniwersytetu Harvarda – zarówno studenci, jak i pracownicy tej uczelni – mają ponadto dostęp do funduszu Harvard Open Access Publishing Equity (HOPE). Zgromadzone w nim środki służą pokryciu kosztów publikacji w otwartych czasopismach w maksymalnej wysokości 3000 USD rocznie na osobę. Kwota ta może zostać spożytkowana na pokrycie kosztów publikacji jednego lub kilku artykułów. W przypadku artykułów posiadających wielu autorów powinni oni podzielić koszt publikacji w otwartym czasopiśmie po równo.

12 Zob E. Priest, *Copyright and the Harvard Open Access Mandate*, „Northwestern Journal of Technology and Intellectual Property”, Vol. 10, 2012. W tym samym artykule autor przedstawia argumentację na rzecz tezy, iż wymóg, by pracownicy uczelni dodatkowo potwierdzali postanowienia mandatu poprzez pisemne udzielenie licencji, w rzeczywistości nie jest niezbędny do tego, by zobowiązanie takie było skuteczne.

13 Dane z dnia 28 grudnia 2015 r. za stroną <https://osc.hul.harvard.edu/dash/mydash>.

Ze środków funduszu mogą być pokryte jedynie koszty publikacji w czasopismach otwartych, a nie hybrydowych. Czasopismo takie musi być ponadto uznanym otwartym czasopismem: wymaga się od niego indeksacji w bazie Directory of Open Access Journals (DOAJ), a jego wydawca musi być członkiem organizacji Open Access Scholarly Publishers Association (OASPA) albo stosować się do jej wytycznych. Wymagania takie powodują, że środki z funduszu nie wspierają czasopism wydawanych w modelu hybrydowym, który bywa krytykowany za tzw. *double dipping*, czyli wielokrotne pobieranie opłat za ten sam artykuł (raz po stronie autora, a następnie po stronie czytelnika) oraz tzw. drapieżnych periodyków (*predatory journals*), tj. niskiej jakości czasopism pozorujących jedynie procedury recenzyjne i nastawionych wyłącznie na maksymalizację zysku, którego źródłem są opłaty po stronie autora.

### 1.3. Uniwersytet w Liège

Polityka otwartego dostępu przyjęta w maju 2007 roku przez Uniwersytet w Liège (jedną z trzech najważniejszych francuskojęzycznych uczelni w Belgii) obowiązuje na tej uczelni od listopada 2008 roku. Weszła ona w życie w wyniku decyzji administracyjnej podjętej za sugestią rektora przez Radę Administracyjną (*Administrative Board*) Uniwersytetu, a nie głosowania pracowników naukowych uczelni lub jej poszczególnych wydziałów.

Stanowi ona przykład polityki typu *immediate deposit/optional access* (natychmiastowy depozyt – opcjonalny dostęp). Pracownicy Uniwersytetu w Liège zostali zobowiązani do wprowadzenia do uczelnianego systemu ORBi danych bibliograficznych wszystkich swoich publikacji naukowych wydanych po roku 2002, a w przypadku artykułów naukowych – również ich pełnych tekstów. System ORBi pełni więc dwojaką rolę: jest to zarówno bibliografia instytucjonalna, jak i repozytorium. Repozytorium jest przy tym otwarte na pełne teksty publikacji innych typów: książki, ich rozdziały czy materiały konferencyjne. W tym przypadku uczelnia nie nakłada jednak na autorów obowiązku analogicznego do tego, który dotyczy artykułów naukowych.

Tam, gdzie jest to możliwe (np. wydawca wyraża na to ogólną zgodę), od autorów oczekuje się, że zapewnią otwarty dostęp do deponowanych przez siebie publikacji. Jeśli w danym przypadku wydawca nie zgadza się na udostępnienie publikacji w sposób otwarty lub będzie to możliwe dopiero po upływie okresu embarga, dostęp do tekstu można uzyskać poprzez funkcję repozytorium pozwalającą wysłać do autora odpowiednią prośbę. W ten sposób, choć sam tekst pozostaje formalnie zamknięty, uzyskanie dostępu do niego wymaga jedynie wykonania kilku prostych czynności przez osobę deponującą oraz użytkownika repozytorium.

Deponować publikacje w repozytorium ORBi mogą jedynie autorzy tych publikacji lub przedstawiciele autorów. W obrębie repozytorium każdy pracownik może wskazać tylko jedną osobę pełniącą tę funkcję (np. swojego asystenta czy doktoranta). Ostateczne zatwierdzenie rekordu zawsze należy jednak do samego autora. Rozwiązanie to ma na celu zaangażowanie pracowników naukowych w upowszechnianie tworzonych przez nich treści, co pozwala na zerwanie z przekonaniem, iż rola naukowca kończy się w momencie przyjęcia przez wydawcę tekstu do publikacji. To m.in. dzięki zastosowaniu takiego rozwiązania w repozytorium uczelni udało się zgromadzić również wiele materiałów, wobec których nie wprowadzo-

no administracyjnego obowiązku rejestracji czy depozytu. W styczniu 2013 roku w bazie znajdowało się 90 tys. rekordów bibliograficznych, z czego ponad 55 tys. posiadało pełne teksty. W przypadku objętych mandatem artykułów wydanych w 2002 roku lub później pełne teksty posiada 100% rekordów (system repozytoryjny nie pozwala tu na wprowadzenie metadanych pozbawionych pełnego tekstu). Jeśli chodzi o inne materiały wydane w tym samym okresie, odsetek ten wynosił 49,5%, a w przypadku materiałów wszystkich typów wydanych przed 2002 rokiem – nieco ponad 40%<sup>14</sup>. Około połowa pełnych tekstów publikacji umieszczonych w repozytorium jest dostępna w modelu otwartym<sup>15</sup>.

Repozytorium działa od listopada 2008 roku w oparciu o oprogramowanie DSpace. Za jego prowadzenie oraz szeroko rozumiane wsparcie użytkowników odpowiada biblioteka uniwersytecka. Pracownicy naukowci uczelni mogą z jej strony liczyć tak na wsparcie o charakterze technicznym, jak i prawnym (zarówno w postaci szeregu przygotowanych wcześniej dokumentów, takich jak FAQ czy załączniki autorskie, jak i osobistego wsparcia ze strony prawnika). Obok repozytorium ORBi na Uniwersytecie w Liège funkcjonują też równoległe repozytorium BICTEL/e, przechowujące pełne teksty rozpraw doktorskich broniących na tej uczelni, oraz platforma PoPuPs, służąca upowszechnianiu treści publikowanych w czasopiśmie wydawanych przez Uniwersytet w sposób otwarty.

Modyfikacje wprowadzone w Liège do systemu DSpace pozwalają na uniknięcie konieczności dublowania pracy w przypadku tekstów wieloautorskich oraz kolektywną korektę ewentualnych błędów przez wszystkich autorów publikacji oraz redaktorów systemu po stronie biblioteki.

Na tle innych polityk instytucjonalnych politykę otwartego dostępu Uniwersytetu w Liège wyróżnia to, że już kilka miesięcy po jej wprowadzeniu wszystkie decyzje dotyczące ewaluacji, zatrudnień, awansów czy alokacji środków budżetowych dokonywane były wyłącznie w oparciu o dane bibliograficzne znajdujące się w repozytorium. Należy przy tym zaznaczyć, iż zasada ta dotyczy danych bibliograficznych, a nie pełnych tekstów publikacji czy też tekstów, do których udało się skutecznie zapewnić otwarty dostęp za pomocą repozytorium instytucjonalnego Uniwersytetu w Liège. W praktyce znaczy to, iż pracownicy Uniwersytetu w Liège otrzymują silny bodziec do tego, by umieszczać w repozytorium metadane swoich publikacji (a tam, gdzie to możliwe, również pełne teksty, najlepiej w otwartym dostępie), jednocześnie jednak nie są poszkodowani w wyniku działań podejmowanych przez osoby i instytucje niechętne otwartemu dostępowi (np. wydawnictwa *a priori* sprzeciwiające się deponowaniu tekstów publikacji w repozytorium instytucjonalnym lub stosujące bardzo długie okresy embargo).

#### 1.4. Węgierska Akademia Nauk

Otwarty mandat Węgierskiej Akademii Nauk (MTA) ukazał się we wrześniu 2012 roku. Jego postanowienia weszły w życie 1 stycznia 2013 roku i stosują się do wszystkich publikacji (artykułów, książek oraz rozdziałów w książkach) złożonych do publikacji po tej dacie. W przypadku tych, które zgłoszono wcześniej, oraz tych,

14 Dane za: [http://orbi.ulg.ac.be/bitstream/2268/150854/1/BernardRentier\\_EuropeanWorkshopUMinho.pdf](http://orbi.ulg.ac.be/bitstream/2268/150854/1/BernardRentier_EuropeanWorkshopUMinho.pdf).

15 <http://initiatives.exlibrisgroup.com/2013/05/a-mandate-for-open-access-university-of.html>



które już się ukazały, postanowienia mandatu przyjmują słabszą formę zaleceń. Podobnie jak w przypadku Uniwersytetu w Liège, o przyjęciu mandatu w MTA zdecydowały władze instytucji, a dokładnie jej prezes.

Zgodnie z treścią otwartego mandatu jego postanowienia uznaje się za zrealizowane zarówno w przypadku obrania zielonej, jak i złotej drogi otwartego dostępu. W przypadku zielonej drogi autorzy winni skorzystać z repozytorium REAL Węgierskiej Akademii Nauk, choć dopuszczalna jest również archiwizacja w repozytorium innej instytucji lub w repozytorium dziedzinowym (np. ArXiv). W przypadku złotej drogi możliwa jest zarówno publikacja w czasopiśmie typu otwartego, jak i w czasopiśmie hybrydowym. Mandat dopuszcza archiwizację ostatecznych wersji autorskich publikacji. W przypadku artykułów dopuszcza się okres embargo, przy czym załącznik autorski sugeruje, by był on nie dłuższy niż jeden rok. W przypadku publikacji, których wydawcy nie zgadzają się na żadną formę otwartego dostępu, autorzy wciąż zobowiązani są do ich zdeponowania w repozytorium REAL, ale ogólnodostępne są jedynie ich metadane.

Otwarty katalog serwisów, w których autorzy objęci otwartym mandatem mogą w zgodzie z nim deponować swoje publikacje, powoduje, że pełne teksty publikacji stanowiących dorobek naukowy Akademii mogą znajdować się w bardzo różnych lokalizacjach. Z tego powodu metadane publikacji oraz URL jej pełnego tekstu winny zostać równolegle wprowadzone do prowadzonej przez Węgierską Akademię Nauk bazy MTMT, która pełni rolę bazy bibliograficznej, indeksu cytowań oraz swoistego agregatora informacji o dorobku publikacyjnym węgierskich naukowców oraz lokalizacji pełnych tekstów. Dzięki takiemu rozwiązaniu możliwy jest monitoring pozwalający określić efektywność obowiązującego mandatu (i tak, w roku 2013 w otwartym dostępie znalazło się ok. 10% dorobku publikacyjnego MTA) i w ślad za tym odpowiednio pokierować inicjatywami o charakterze szkoleniowym czy porządkowym.

Autorzy objęci mandatem otrzymują ponadto materiały ułatwiające im negocjacje z wydawcami w postaci wzoru załącznika autorskiego do umowy oraz listu przewodniego informującego o zasadach dotyczących otwartości obowiązujących w MTA. W przypadku gdy załącznik autorski (dostępny w języku angielskim) zostanie zaakceptowany przez wydawcę, autor publikacji zachowuje prawo do jej zdeponowania w repozytorium instytucjonalnym lub dziedzinowym (w postaci ostatecznej wersji autorskiej lub – jeśli pozwala na to umowa – ostatecznej wersji wydawcy) w momencie jej opublikowania. Ma również prawo otwarcia do niej dostępu nie później niż w ciągu 12 miesięcy od momentu opublikowania, a w przypadku publikacji udostępnionych w modelu otwartym przez wydawcę – natychmiastowo.

## 1.5. Politechnika Federalna w Zurychu

Otwarty mandat został przyjęty przez Politechnikę Federalną w Zurychu w lipcu 2005 roku, zaś w życie wszedł 1 lipca 2008 roku. Również w tym przypadku przyjęcie mandatu miało charakter decyzji władz, tym razem podjętej przez zarząd (*Schulleitung*) tej uczelni.

Mandatem objęci są zarówno pracownicy naukowcy, jak i studenci studiów magisterskich i doktoranckich

(*post-graduate students*) Politechniki, od których wymaga się, by możliwie jak najszybciej (dokument nie precyzuje, w którym konkretnie momencie ma to nastąpić) udostępniali w sposób otwarty swoje publikacje naukowe: artykuły, książki i ich rozdziały, a także prace magisterskie i rozprawy doktorskie. Serwisem służącym upowszechnianiu tych treści jest instytucjonalne repozytorium ETH E-Collection, prowadzone przez tamtejszą bibliotekę w oparciu o oprogramowanie repozytoryjne Fez.

Autorzy objęci mandatem są ponadto zachęceni do tego, by publikować swoje teksty w otwartych czasopismach. Politechnika w Zurychu ułatwia taki sposób postępowania, zawierając umowy z wydawcami takich periodyków (m.in. PLOS oraz Springer), dzięki czemu z perspektywy pracowników i studentów uczelni publikacja w nich jest bezpłatna. Uczelnia nie pokrywa natomiast kosztów publikacji w czasopismach hybrydowych<sup>16</sup>, choć część czasopism tego rodzaju prenumerowanych przez miejscową bibliotekę oferuje możliwość publikacji po obniżonych kosztach<sup>17</sup>.

O wyjątkowości mandatu Politechniki w Zurychu stanowi ponadto, iż uczelnia oczekuje, by autorzy – w tych przypadkach, w których jest to możliwe – zachowywali prawa autorskie do swoich tekstów. Rolę punktu wspierającego osoby objęte postanowieniami mandatu Politechniki pełni biblioteka tej instytucji. Na wsparcie takie składają się m.in. dotyczące otwartości kursy i szkolenia<sup>18</sup> czy też pomoc prawna udzielana np. w razie wątpliwości dotyczących legalności udostępnienia danego tekstu w modelu otwartym. W repozytorium Politechniki w Zurychu znajduje się obecnie po ok. 1000–1200 dokumentów wydanych w każdym roku z przedziału 2008–2014.

## 1.6. Podsumowanie

Przedstawione powyżej przykłady otwartych mandatów różnią się, jeśli chodzi o ich skuteczność oraz mechanizmy wykorzystywane do osiągnięcia zakładanych celów. Idący pod tym względem najdalej mandat Uniwersytetu Harvarda zmienia dotychczasową perspektywę publikowania w tym sensie, że model otwarty czyni modelem domyślnym; to odstępstwo od niego wymaga podjęcia przez autora dodatkowych działań w postaci złożenia prośby o tzw. *waiver*. Mandat ten jest jednak silnie zakorzeniony w amerykańskim prawie autorskim oraz orzecznictwie tamtejszych sądów, co sprawia, że wykorzystanie takiego samego mechanizmu w obrębie innych jurysdykcji może nie być możliwe.

Zupełnie inny mechanizm wykorzystuje mandat Uniwersytetu w Liège, który w odniesieniu do pozycji wydanych w 2002 roku i później nakłada jedynie obowiązek deponowania pełnych tekstów (ale nie zapewnienia do nich otwartego dostępu) w przypadku artykułów naukowych oraz wprowadzenia metadanych (ale nie dokonania depozytu, a tym bardziej zapewnienia otwartego dostępu) w przypadku

---

16 Jest to rozwiązanie odwrotne od tego, które funkcjonuje w Polsce, gdzie naukowcy mają do swojej dyspozycji program Springer Open Choice, obejmujący wyłącznie czasopisma hybrydowe wydawnictwa Springer, ale już nie czasopisma w pełni otwarte.

17 Więcej informacji na temat możliwości publikowania przez autorów afiliowanych w ETH Zurich w czasopismach otwartych i hybrydowych znaleźć można na stronie internetowej: <https://www.library.ethz.ch/en/ms/Open-Access-at-ETH-Zurich/Publishing-in-open-access-journals/Publishing-in-open-access-journals-Funding>.

18 Ich zestawienie znaleźć można na stronie: <https://www.library.ethz.ch/en/ms/Open-Access-at-ETH-Zurich/Consultation-courses>.

publikacji innych typów. Dzięki licznym działaniom towarzyszącym, takim jak decyzja o uwzględnianiu do celów ewaluacyjnych i administracyjnych jedynie danych znajdujących się w repozytorium, silne zaangażowanie autorów w proces dostarczania danych i późniejsze nimi zarządzanie oraz udostępnienie pracownikom szerokich statystyk, belgijskiej uczelni udało się sprawić, iż nałożone obowiązki realizowane są z nawiązką, a znaczna część materiałów znajdujących się w jej repozytorium dostępna jest w sposób otwarty. Wskazuje to, iż wprowadzenie nawet ograniczonego zobowiązania – dzięki odpowiedniemu przepływowi danych, zapewnieniu wsparcia oraz przemyślanemu systemowi zachęt i kar – może skutkować lepszą dostępnością dużej części dorobku danej instytucji.

# Rozdział 2

## Kontekst polski – przekonania, praktyki, oczekiwania

W ramach prac nad niniejszym raportem zrealizowany został zogniskowany wywiad grupowy (FGI), w którym wzięło udział osiem osób zawodowo związanych z tworzeniem oraz upowszechnianiem wiedzy naukowej. Dwoje uczestników stanowili naukowcy: badaczka posiadająca stopień doktora habilitowanego, reprezentująca obszar nauk humanistycznych, oraz naukowiec posiadający tytuł profesora, pełniący obecnie funkcję kierowniczą w obrębie swojej jednostki, reprezentujący obszar nauk ścisłych.

W badaniu uczestniczyli ponadto przedstawiciele środowiska bibliotekarzy naukowych, reprezentujący zarówno uczelnie publiczne, jak i niepubliczne uczelnie akademickie. Wśród tej grupy uczestników znalazły się osoby piastujące funkcje kierownicze, posiadające doświadczenie w prowadzeniu repozytoriów instytucjonalnych i zarządzaniu elektronicznymi bazami danych. Grono uczestników uzupełniali przedstawiciele wydawnictw naukowych – wydawnictwa uczelnianego oraz wydawnictwa komercyjnego publikującego częściowo w modelu otwartym – oraz przedstawiciel redakcji humanistycznego czasopisma naukowego wydawanego w modelu tradycyjnym, które zapewnia jednak opóźniony otwarty dostęp do publikowanych w nim artykułów.

Wszystkie osoby biorące udział w FGI pracują na co dzień w instytucjach zlokalizowanych w Warszawie. Samo spotkanie także odbyło się w Warszawie, a poprowadził je zewnętrzny moderator, niezwiązany z ruchem otwartego dostępu. Wywiad rejestrowany był za pomocą kamery, a z powstałego zapisu sporządzona została transkrypcja. Cytaty z wypowiedzi uczestników spotkania znajdujące się w niniejszym raporcie w celu ułatwienia odbioru tekstu zostały poddane redakcji. Autorzy dołożyli przy tym starań, by wprowadzane zmiany nie naruszały ich pierwotnego sensu.

Scenariusz FGI obejmował dwa bloki pytań. Pierwszy z nich poświęcony był formie publikacji naukowych i pełnionej przez nie obecnie roli. Drugi dotyczył kształtu oraz znaczenia otwartych modeli publikacyjnych oraz ewentualnych trudności i nieporozumień, jakie się z nimi wiążą.

### 2.1. Forma i rola publikacji naukowych

Uczestnicy wyraźnie wskazywali na istotne różnice pomiędzy dominującymi formami publikowania w poszczególnych obszarach wiedzy. O ile w przypadku nauk humanistycznych kluczową formą publikacji

pozostaje monografia, o tyle w przypadku nauk ścisłych najistotniejszą rolę pełnią artykuły naukowe. Przedstawicielka nauk humanistycznych dostrzegała jednak rosnącą rolę artykułów również w swoim obszarze wiedzy, za co odpowiadać mają obowiązujące zasady oceny jednostek naukowych i samych naukowców:

OS1: [...] *artykuły [w naukach humanistycznych – przyp. aut.] są drugorzędne, chociaż zmienia się to ze względu na zmiany systemowe, to znaczy narzuca się w pewnym sensie publikacje artykułów [...], aczkolwiek główny nacisk kładziony jest na monografie.*

Uczestniczka zaznaczyła jednocześnie, iż obecnie monografia nie musi przyjmować już tradycyjnej drukowanej postaci, choć nośnik papierowy wciąż jest pozytywnie wartościowany przez dużą część odbiorców publikacji naukowych. Przekonanie o większej wartości papierowych monografii i artykułów naukowych jest przy tym wyłącznie kwestią nawyków i przyzwyczajzeń. Dzieje się tak bowiem mimo braku prawnych wymagań dotyczących koniecznego czy też oczekiwanego nośnika publikacji:

OS1: [...] *Bardzo wolno przebija się do świadomości powszechnej, że to jest równoważne. W wielu dyscyplinach to nie robi różnicy jakościowej. Co gorsza, ten problem dotyczy także tych, którzy oceniają naukowców. Jest to bardzo wyraźne w praktykach recenzyjnych, na przykład dorobku naukowego. W dyscyplinach takich, jak dyscypliny humanistyczne i społeczne, kiedy się przedstawia dorobek... Tego nie ma nigdzie na papierze. Nie ma tego w przepisach. Ale tradycja, uzus, praktyka, zwyczaj jest taki, że egzemplarz drukowany być powinien. To jest absurdalne, ale tak właśnie jest.*

Przekonanie o większej wartości publikacji papierowych jest więc podzielane również przez osoby mające decydujący wpływ na ocenę dorobku i – co za tym idzie – rozwój kariery naukowej tych autorów, którzy sami z siebie nie są w szczególności przywiązani do tradycyjnego nośnika i nie sądzą, żeby publikacje papierowe miały większą wartość. W konsekwencji jednak również ci autorzy – niejako z konieczności – decydują się na publikację z wykorzystaniem tradycyjnego nośnika, chcąc sprostać oczekiwaniom osób, od których zależą ich dalsze zawodowe losy.

Również przedstawiciel nauk ścisłych stwierdził, iż nośnik papierowy – choć w mniejszym stopniu – jest pozytywnie wartościowany w jego obszarze wiedzy. Także tu przeświadczenie to ma charakter nieformalny i nie posiada ugruntowania w istniejących przepisach regulujących np. zasady oceny jednostek naukowych, pracowników naukowych czy przyznawania stopni naukowych i tytułu naukowego.

Przedstawicielka wydawnictwa komercyjnego wydającego książki i czasopisma (w tym także w modelu otwartym) zwróciła ponadto uwagę na to, że choć z jednej strony posiadanie – postrzeganego jako prestiżowy – egzemplarza papierowego samo w sobie stanowi nagrodę za nierzadko wieloletni wysiłek twórczy, z drugiej autorzy zdają sobie sprawę, iż za pomocą darmowej elektronicznej kopii będzie im dużo łatwiej dotrzeć do czytelników pozostających poza zasięgiem tradycyjnych modeli dystrybucji treści naukowych:

OS6: *Autorzy [...] chcą mieć tę papierową wersję, dłużej pracują nad tym. Jak ktoś pracuje nad książką pięć lat, to chce to widzieć na...*

Moderator: *To chce mieć ją na półce.*

OS6: *Ale też z drugiej strony, jak autorzy [...] piszą na jakiś wąski, niszowy temat w Polsce albo gdzie indziej, to chcą, żeby ta książka była wszędzie. Żeby naukowiec, który studiuje robotykę w Chinach, też to przeczytał, nie tylko te małe grona tutaj w Polsce. I to też staje się coraz ważniejsze.*

Jedna z przedstawicielek środowiska bibliotekarskiego zwróciła natomiast uwagę na to, że dla osób korzystających z usług bibliotek ogromne znaczenie ma łatwość dotarcia do informacji oraz fakt, iż dociera się dokładnie do tego jej fragmentu, który jest użytkownikowi potrzebny. W tym sensie – z punktu widzenia osób związanych z biblioteką – dobra publikacja to taka, którą łatwo jest odnaleźć i z której łatwo jest wyodrębnić fragmenty w danym kontekście istotne dla czytelnika:

OS5: *Ale student studiów pierwszego i drugiego stopnia chce, żeby ta informacja była najświeższa i często jest też tak, że on nie potrzebuje całej publikacji, tylko na przykład to, co polecił wykładowca. I ja proponuję publikacje w wolnym dostępie<sup>19</sup>, zapraszam tego pana do dziedziny takiej i takiej, w której znajdzie piętnaście publikacji na ten temat. Ale on chce tylko ten jeden rozdział. I to jest jakby kwestia poziomu wtajemniczenia w poszukiwanie informacji i zgłębienia jakichś zagadnień.*

Oczekiwania tego typu żywiłyby więc przede wszystkim osoby posiadające niskie kompetencje informacyjne, których poziom utrudnia im samodzielne odnalezienie np. konkretnego fragmentu tekstu zadanego przez osobę prowadzącą zajęcia. Wskazuje to, iż większego znaczenia niż wyrażane wprost przez uczestników FGI nabierałaby kwestia wolnych licencji, pozwalających na praktycznie dowolne tworzenie antologii tekstów czy zestawów materiałów dydaktycznych, oraz kwestia opatrzenia udostępnianych zasobów odpowiednim zestawem metadanych, umożliwiającym łatwy dostęp do nich przez ogólne i specjalistyczne wyszukiwarki internetowe.

Dyskusja między uczestnikami FGI wskazuje również, że niezagrożona pozostaje istotna rola czasopism naukowych rozumianych jako instytucje, których zadaniem jest dbanie o jakość publikowanych treści i które same w sobie są jej probierzem. Wyznacznikiem tego, jak dobrze czasopismo sprawdza się w tej roli – jak zauważył uczestnik reprezentujący redakcję czasopisma humanistycznego – jest przede wszystkim lista osób z nim związanych (rada redakcyjna, redakcja czy recenzenci), jakość recenzji oraz – w szczególności w humanistyce – wspomniany już nośnik papierowy. Co więcej, jak stwierdził przedstawiciel nauk ścisłych, to właśnie przez pryzmat czasopisma nierzadko postrzegana jest też pozycja samego badacza:

OS8: *[...] Wydaje mi się, że na [...] rynku czasopisma spełniają taką funkcję, że jakość czasopisma, jego środowiskowy obraz [...] bardzo wpływa na to, jak ktoś podczas swojej kariery naukowej jest postrzegany przez otoczenie. Jak szybko awansuje, jak łatwo mu o stypendia, pozycje, kontakty, akceptację swoich osiągnięć. To jest, moim zdaniem, absolutnie podstawowa rola czasopism.*

---

19 Tj. w tej części biblioteki, która umożliwia czytelnikom bezpośredni dostęp do tradycyjnych, drukowanych wydań.

Uczestnicy rozmaicie oceniali z kolei wartość miar ilościowych, takich jak Impact Factor, interpretowanych często jako odzwierciedlenie jakości publikowanych w danym czasopiśmie treści. Z jednej strony wskazywano na nieporównywalność nominalnych wartości wskaźnika IF pomiędzy dyscyplinami, jego relatywną podatność na manipulację przez „spółdzielnie” wzajemnie cytujących się autorów czy fakt, iż nie stanowi on prostego wskaźnika jakości pojedynczej publikacji ukazującej się w danym periodyku. Z drugiej jednak strony przedstawicielki środowiska bibliotecznego zwracały uwagę, że – przy wszystkich tych zastrzeżeniach – miary ilościowe są istotnym parametrem opisującym poczytność czy też popularność i jako takie mogą również być ważne dla odbiorców:

*OS7: Ale popularność jest dosyć istotna. [...] Przecież są takie narzędzia, które śledzą ruch, też taki w mediach społecznościowych, i które mówią [...] o tym, kto cytuje prace, kto się powołuje, ile jest pobrań z jakiego miejsca. To narzędzia, które są w stanie to zdefiniować i też mogą mieć jakąś wiedzę o tej publikacji. Czyli niejako mogą wiedzieć, czy to jest istotne, czy nieistotne.*

Mówiąc o różnicach między publikacjami elektronicznymi a papierowymi, przedstawiciel uczelnianego wydawnictwa naukowego zwrócił z kolei uwagę na pogarszającą się jego zdaniem jakość edytorską publikacji naukowych. „Ostateczność” wydań papierowych miałyby tu niejako wymuszać większą staranność, która nie jest konieczna w dobie łatwych do późniejszej korekty publikacji elektronicznych:

*OS4: [...] Wszyscy chcą stosować rachunek ekonomiczny. Wydanie elektronicznych publikacji jest dużo tańsze niż tradycyjne, papierowe, i ta jakość idzie w kierunku różnym, niekoniecznie bardzo dobrym czy polepszania tej jakości. Wiadomo, że czeka nas wszystkich coraz więcej publikowania tych materiałów naukowych w wersji elektronicznej, do tego trzeba się przygotować.*

Ponadto uczestnik zwrócił uwagę na rolę ewaluacji jednostek i czasopism, których zasady kształtowane są przede wszystkim przez Ministerstwo Nauki i Szkolnictwa Wyższego. Zasady te są nie tylko narzędziami służącymi obiektywizacji pomiaru jakości pracy naukowej poszczególnych osób i instytucji, ale i sposobem kształtowania polityki naukowej państwa, wpływającym na faktyczne zachowania redakcji czasopism, instytucji naukowych i samych naukowców (uwagi o podobnej treści pojawiały się kilkakrotnie również w wypowiedziach innych uczestników wywiadu).

## **2.2. Forma i funkcje otwartego dostępu**

Próbując określić minimalne warunki, jakie musi spełniać publikacja dostępna w modelu otwartym, rozmówcy akcentowali kwestię braku opłaty po stronie końcowego użytkownika, niezależnie od tego, czy jest on związany z instytucją naukową, czy nie. Przedstawiciel czasopisma humanistycznego zwrócił przy tym uwagę na to, iż „opłatę” w przypadku dostępu do materiałów znajdujących się w Internecie należy rozumieć szeroko. Może mieć ona charakter pieniężny, ale może też przyjmować postać informacji, jakich czytelnik chcący uzyskać dostęp do treści musi udzielić na własny temat:

OS3: *Płaci się danymi osobowymi, tak samo jak na Facebooku. Płaci się swoim adresem mailowym, płaci się tym, że można sobie wejść, że dostaje się newsy i tak dalej.*

Wśród cech, jakie powinna posiadać publikacja w otwartym dostępie, nie wymieniano natomiast wolnych licencji, których znajomość i zrozumienie w społeczności akademickiej oceniane były skądinąd jako bardzo niskie. Otwarty dostęp w rozumieniu uczestników wywiadu był więc bliższy jego wersji gratis niż wersji libre (tj. uwzględniającej możliwość bardzo szerokiego wykorzystania tekstów).

Przedstawicielka środowiska bibliotekarskiego zaznaczyła ponadto, iż o otwartym dostępie można mówić w jej ocenie jedynie wtedy, gdy poszanowane są prawa autorskie podmiotów będących ich posiadaczami. Miałyby to odróżniać teksty dostępne w modelu otwartym np. od publikacji udostępnianych niekiedy z naruszeniem praw autorskich w popularnych serwisach hostingowych:

OS2: *[Ż]eby jeszcze dodać, że się nie łamie praw autorskich. Bo jak usłyszałam, że [nazwa serwisu hostingowego – przyp. aut. raportu] to jest otwarty dostęp, to... Okej, w takiej funkcji, w ideologii – być może. Natomiast w takiej funkcji praktyczno-prawnej, to już mi się zjeżyło wszystko...*

Wśród okoliczności przemawiających za otwartym dostępem uczestnicy wymieniali fakt finansowania badań naukowych ze środków publicznych oraz wysokie koszty dostępu do artykułów publikowanych w czasopiśmie wydawanych przez podmioty komercyjne, widoczne zwłaszcza w zestawieniu z dużo niższymi opłatami, jakie trzeba ponieść, by uzyskać dostęp do dobrej jakości tytułów wydawanych przez towarzystwa naukowe i uniwersytety. Tymczasem sami autorzy nie czerpią żadnych korzyści finansowych ze sprzedaży swoich utworów, co naukowiec z obszaru nauk ścisłych opisał w następujących słowach:

OS8: *Oni żyją z naszej krwi. My to wszystko piszemy za darmo, recenzujemy za darmo, za darmo dajemy nasze nazwiska do czasopism, jeszcze sami to zwykle redagujemy i składamy w profesjonalny skład, a oni to potem wrzucają w swój Internet, gdzie jeden gigabajt przestrzeni dyskowej jest wiadomo jak tani, i sprzedają to za grube tysiące dolarów czy euro. Zbundlowane paczki<sup>20</sup>.*

Przedstawicielka nauk humanistycznych, odnosząc się do kwestii bardzo wysokich cen publikacji naukowych – w tym książek – zwróciła uwagę, że grozi to tym, iż pewne treści w ogóle nie dotrą do odbiorców. Sytuacja taka jest tymczasem niepożądana zarówno z perspektywy czytelników, jak i autorów:

OS1: *Jeżeli mam publikacje monograficzne, humanistyczne, których cena tomu, niedużej książeczki, trzystustronicowej wynosi, dajmy na to, dwieście euro, to to jest cena wysoka nawet poza Polską. A w Polsce w ogóle zaporowa dla uczelni. Kiedy publikacja będzie dostępna w modelu otwartym, automatycznie będzie miała większy zasięg.*

W odniesieniu do odbiorców spoza świata akademickiego uczestniczka zwróciła ponadto uwagę na istotny

---

20 Chodzi tu o transakcje dotyczące dużych zestawów czasopism, zawierane z wydawcami komercyjnymi (*ang. big deals*).



jej zdaniem aspekt społeczny, związany z koniecznością zapewnienia transparentności rezultatów badań naukowych, wynikającą z faktu ich finansowania ze środków publicznych. Ta sama rozmówczyni zaznaczyła jednocześnie, że naukowcy chcący publikować w modelu otwartym bywają zmuszeni do dokonywania wyboru pomiędzy publikacją u bardziej prestiżowego, tradycyjnego wydawcy, który czerpie zyski ze sprzedaży egzemplarzy utworu naukowego, a wydawcami dającymi możliwość publikacji tekstu w modelu otwartym, którzy jednak nie mogą poszczycić się równie dużym poważaniem w środowisku naukowym. Kluczowy jest przy tym nie nośnik, a model dystrybucji – prestiżowi wydawcy tradycyjni mogą sprzedawać elektroniczne wersje książki, podczas gdy wydawcy otwarci udostępniają je bezpłatnie:

*OS1: Mówię [...] o sytuacji trudnego wyboru między publikacją w otwartym, dostępnym trybie, ale poprzez wydawnictwo działające w tym modelu. Czyli zakładamy, że wybieramy między wydawnictwem, które jako wydawnictwo openaccessowe jednak nie ma tego prestiżu, co stare wydawnictwa drukowane, ale mamy świadomość, że ta publikacja będzie udostępniona szeroko, a publikacją w zamkniętym wydawnictwie, które nas pozbawia praw, bo jeszcze każe nam podpisać umowę, która jest niekorzystna, na wyłączność. [...] Tu nie chodzi o papier, tu chodzi o prestiż. Bo oni mogą to mieć też w wersji ebookowej. Bo to już nie jest ważne, jaki jest nośnik. Ja już na samym początku powiedziałam. W tym momencie nie jest ważny nośnik. Ważny jest ten model publikacyjny. Czyli to jest wybór, przed którym naukowiec stoi. To jest dość trudny wybór.*

W skrajnym przypadku naukowiec może więc stanąć przed wyborem między publikacją papierową, wydaną i dystrybuowaną w modelu tradycyjnym przez uznane, prestiżowe wydawnictwo naukowe, a publikacją elektroniczną w relatywnie nowym, otwartym wydawnictwie, oferującym wprawdzie bardzo szeroką dostępność tekstu, ale niemogącym poszczycić się porównywalnym prestiżem. Tym dwóm rozwiązaniom towarzyszyć może całe spektrum rozwiązań pośrednich, takich jak dostępne za opłatą elektroniczne publikacje w ofercie uznanych wydawnictw, papierowe pozycje dostępne za opłatą i wydawane przez relatywnie nowe podmioty czy elektroniczne wersje otwarte współistniejące z płatnymi wydaniem papierowymi.

Decydując się na wybór miejsca publikacji, autor musi więc uwzględnić co najmniej trzy czynniki: sposób dystrybucji, nośnik oraz prestiż wydawnictwa stanowiący kryterium postrzegania jakości jego utworu. Trudność decyzji, na którą wskazuje uczestniczka wywiadu, wynika z tego, że prestiż idzie często w parze z odpłatnością po stronie czytelnika (a można dodać, że nierzadko również z dostępnością jedynie wersji papierowej, która skądinąd sama w sobie bywa ceniona wyżej), zaś otwarta wersja elektroniczna – z relatywnie niższym prestiżem wydawcy. Należy jednak zauważyć, iż związek taki ma charakter przygodny i co do zasady nic nie stoi na przeszkodzie, by uznane wydawnictwa rozszerzyły swą ofertę o wydania otwarte lub w całości zmieniły swój model dystrybucji, a nowe wydawnictwa otwarte z czasem zyskiwały porównywalny prestiż.

Uczestnicy sygnalizowali ponadto, że otwarty dostęp jest optymalnym modelem w przypadku czasopism o bardzo niskich nakładach, publikujących artykuły z wąskich dziedzin, w których przypadku koszty druku sprawiają, że tradycyjny model dystrybucji przestaje się bilansować:

OS2: *Można już mówić o pewnych czasopismach, które są tak wąskotematyczne, że inaczej niż jako open acces nie są wydawane. Po prostu koszty wydruku tego czasopisma się nie zwrócą. [...] Parę osób to przeczyta, ale dla rozwoju tej konkretnej dziedziny nauki jest to bardzo istotne. [...] I tu nie ma wyboru [...]. Po prostu jest to za wąska dziedzina.*

OS6: *My mamy wiele takich przypadków, takie czasopisma, które tylko mogą istnieć tak, jak panie mówią, że zmieniają model na open access, żeby po prostu dalej działać. Bo nie opłaca się ich drukować.*

Uczestnicy wskazywali również na istniejące napięcie związane z modelem otwartego dostępu bazującym na repozytoriach. Z jednej strony w interesie tradycyjnego wydawcy jest to, by stosować możliwie jak najdłuższe okresy embarga pozwalające mu chronić swoje zyski. Z drugiej dzisiejsze tempo obiegu informacji naukowej jest tak szybkie, iż w niektórych dziedzinach publikacja starzeje się nawet przed upływem relatywnie krótkiego, bo kilkumiesięcznego okresu, w trakcie którego jest ona dostępna wyłącznie za opłatą. Po jego upływie jest już natomiast mało przydatna.

Uczestniczka reprezentująca wydawnictwo komercyjne oferujące możliwość publikacji w otwartym dostępie zwróciła również uwagę na to, iż udostępnianie w sieci wczesnych wersji artykułu stanowi element pracy nad publikacją, a tym samym element badań naukowych. Udostępniony innym badaczom jako *working paper* tekst znajduje krytycznych czytelników jeszcze przed swoją oficjalną publikacją, co pozwala zebrać opinie na jego temat i udoskonalić go jeszcze przed przesłaniem do czasopisma. W rezultacie korzystnie wpływa to na merytoryczną jakość ostatecznej wersji publikacji:

OS6: *To jest taka wersja robocza tego artykułu, taka początkowa. Wrzuca się na arXiv początek jakiejś pracy, ktoś może coś dodać, zmienić jeszcze. Tak się pracuje.*

Uczestniczka ta zauważyła jednocześnie, iż tego typu praktyki popularne są jedynie w niektórych dziedzinach i w odniesieniu do publikacji określonego typu. Dzieje się tak jednak nie dlatego, że w przypadku innych dziedzin i rodzajów publikacji podobne działania nie przyniosłyby analogicznych korzyści, ale dlatego, że komercyjni wydawcy starają się tworzyć bariery mające – w ich ocenie – chronić ich strumień przychodów. Jednym ze sposobów takiej ochrony może być odmowa wydania tekstu opublikowanego uprzednio w Internecie:

OS6: *[A]utor pisze książkę, pisze końcowe rozdziały i puszcza to na swoim blogu. My wtedy tego nie opublikujemy, bo to w całości już gdzieś istnieje na stronie i nie ma dla nas żadnej wartości.*

Zarazem naukowiec z obszaru nauk ścisłych zaznaczył, iż wydawcy powinni pełnić rolę służebną wobec naukowców i to do tych drugich powinien należeć decydujący głos w kwestii tego, w którym momencie i w jaki sposób upowszechniać swoje teksty. Jeśli niektórzy wydawcy starają się w tej kwestii wywierać presję na naukowców, ci ostatni powinni po prostu unikać takich wydawców.

Uczestniczka ze środowiska bibliotekarskiego zwróciła jednak uwagę, iż w podobny sposób postępować mogą przede wszystkim badacze o stosunkowo wysokiej pozycji w środowisku, a więc znajdujący się na dalszych etapach swoich karier. W jej ocenie ci naukowcy, którzy dopiero budują swój prestiż, zgodzą się na dowolne warunki proponowane przez wydawnictwo, gdyż publikacja w uznanym czasopiśmie jest dla nich bardzo atrakcyjna, bez względu na towarzyszące jej ograniczenia wymuszane przez wydawców:

*OS2: Pan mówi z pozycji naukowca, który ma już ugruntowaną pozycję. Natomiast weźmy pod uwagę także młodych doktorantów, którzy chcą publikować i którzy szukają drogi. Oni tak nie krzykną. (...) Jest cały szereg takich dziedzin, w których młody doktorant musi, przynajmniej raz w życiu, trafić do jakiegoś czasopisma uznawanego i on naprawdę przyjmie wszystkie warunki, jakie wydawca mu da. Koniec.*

Przedstawicielka nauk humanistycznych zwróciła uwagę, iż sposobem na wzmocnienie pozycji negocjacyjnej autorów wobec wydawców mogłoby być zobowiązanie do publikacji w otwartym dostępie nałożone np. przez największego i najważniejszego fundatora badań i publikacji naukowych w Polsce:

*OS1: Czyli w tym idealnym świecie, który możemy sobie wyobrazić, ten, kto płaci – czyli w tym wypadku najczęściej w Polsce państwo – narzuca na nas obowiązek publikacji w wolnym dostępie. Wtedy oczywiście mamy pozycję negocjacyjną z wydawcą.*

Ta sama rozmówczyni wskazała również na silne dziedzinowe zróżnicowanie znajomości tematyki otwartego dostępu, a co za tym idzie – zrozumienia zagadnień związanych z zawieraniem umów wydawniczych czy wolnymi licencjami:

*OS1: Natomiast jest drugi problem – bardzo ważny w Polsce – mianowicie problem świadomości. I to jest bardzo różnie w zależności od dyscypliny. Są takie dyscypliny, gdzie w ogóle nie ma dyskusji – model otwarty jest czymś naturalnym. Natomiast są takie dyscypliny, w których wcale nie jest. Dlatego że funkcjonuje cała masa stereotypów i przesądów na temat open acces, w rodzaju: jak ja coś wydam w modelu otwartym, to to jest gorsze z założenia. [...] Czyli jak to jest możliwe, że ja wrzucam coś, co sama wymyśliłam, moja iluśletnia ciężka praca, monografia – ja ją teraz opublikuję ot tak?*

Przedstawiciel nauk ścisłych zwrócił z kolei uwagę na obecne również w jego dyscyplinie przekonanie o ogólnie niskiej jakości publikacji w otwartych periodykach pobierających opłaty za opublikowanie tekstu. W ich przypadku zysk wydawcy wiąże się z samym faktem opublikowania artykułu, co miałyby stanowić zachętę do lekceważenia procesu recenzowania lub wręcz jego pozorowania.

## 2.3. Najważniejsze wnioski w kontekście kształtowania i wdrażania polityki otwartego dostępu

I. Na ocenę jakości publikowanych treści naukowych oraz prestiż czasopism i autorów wpływają zarówno formalne reguły dotyczące na przykład przyznawania stopni i tytułu naukowego czy ewaluacji jednostek i ich pracowników, jak i nieformalne zasady, zwyczaje i przyzwyczajenia środowiska naukowego. Te ostatnie w zależności od dziedziny mogą manifestować się z różną intensywnością i dotyczyć np. nośnika (postrzeganie publikacji papierowych jako bardziej wartościowych), prestiżu czasopisma (widzianego przez pryzmat osób zaangażowanych w jego działalność) czy prestiżu wydawnictwa (większe uznanie dla publikacji wydawanych przez podmioty istniejące na rynku od dłuższego czasu i funkcjonujące najczęściej w tradycyjnym modelu, wymagającym uiszczenia opłaty po stronie czytelnika). Polityka otwartego dostępu winna uwzględniać zarówno czynniki o charakterze formalnym, jak i nieformalnym. Zignorowanie tych drugich może łatwo doprowadzić do nieporozumień dotyczących rzekomych celów czy konsekwencji podejmowanych działań na rzecz otwartości.

II. W ocenie uczestników FGI kwestie związane z otwartym dostępem są w środowisku naukowym ogólnie rzecz biorąc słabo rozumiane. W szczególności dotyczy to zagadnień prawnych, takich jak umowy wydawnicze czy wolne licencje. Istotne różnice pod tym względem rysują się jednak pomiędzy tymi dziedzinami, w których otwarty obieg treści zdążył już okrzepnąć, a tymi, w których stanowi on ciągle relatywną nowość. Działania o charakterze informacyjnym i szkoleniowym powinny więc stanowić integralny element realizacji każdej polityki otwartego dostępu. Ich pominięcie może z kolei utrudnić lub wręcz uniemożliwić jej wdrożenie w wielu dziedzinach.

III. Jednym z najważniejszych celów silnej polityki otwartego dostępu – tj. takiej, która wiąże się z wprowadzeniem obowiązku udostępniania publikacji w sposób otwarty – winno być wzmocnienie pozycji negocjacyjnej naukowców wobec tych wydawców czasopism i książek naukowych, którzy pozostają niechętni otwartemu dostępowi. W obecnej sytuacji – z wyłączeniem kilku dziedzin, w których komunikacja naukowa już dziś oparta jest o duże, międzynarodowe otwarte repozytoria – pozycja pojedynczego autora lub ich niewielkiej grupy pozostaje bardzo słaba. Presja na budowanie dorobku publikacyjnego skłania autora raczej do prostego akceptowania warunków umowy przedstawionej przez wydawnictwo. Dotyczy to przede wszystkim badaczy dopiero budujących swój prestiż i pozycję zawodową, dla których odmowa publikacji przez wydawcę lub jej opóźnienie może stanowić istotną przeszkodę w rozwoju kariery. Przyjęcie bez zastrzeżeń proponowanych postanowień umowy zmniejsza tymczasem powyższe czynniki ryzyka. Przy braku otwartego mandatu (który mógłby wzmocnić pozycję negocjacyjną autora wobec niechętnego otwartemu dostępowi wydawcy), ale i „miękkich” zachęt do publikowania w modelu otwartym (np. stanowiących element systemu ewaluacji dorobku naukowego, a tym samym skłaniających do podjęcia wysiłku w celu zapewnienia otwartego dostępu do tych choćby publikacji, w których przypadku nie jest to trudne czy kosztowne), autorzy ani nie są w bezpośredni sposób motywowani do zabiegania o otwarty dostęp do swoich prac, ani nie otrzymują systemowego wsparcia pozwalającego zwiększyć skuteczność takich zabiegów.

## **II. Otwarty dostęp w polskich instytucjach naukowych – rekomendacje**

# Rozdział 3

## Polityka otwartości

Wdrażanie otwartego dostępu w instytucji naukowej powinno uwzględniać aspekty prawne, organizacyjne, techniczne i komunikacyjne. W tej części raportu prezentujemy najistotniejsze z nich. Rozpoczynamy od zagadnienia polityki otwartości instytucji naukowej, a następnie przechodzimy do spraw związanych z działalnością wydawniczą, otwartymi repozytoriami instytucjonalnymi i działaniami informacyjno-komunikacyjnymi. W każdym wypadku staramy się z jednej strony zarysować kontekst omawianych kwestii, z drugiej zaś przedstawić możliwie szczegółowe rekomendacje ich dotyczące.

Kluczowym elementem dobrej polityki otwartości w instytucji naukowej jest zobowiązanie pracowników naukowych, aby wszystkie swoje publikacje naukowe powstałe w ramach tej instytucji udostępniali w modelu otwartym. W niniejszym rozdziale skupiamy się na sposobach wprowadzenia takiego zobowiązania. Powinno to nastąpić poprzez przyjęcie przez władze danej instytucji naukowej formalnego dokumentu nakładającego omawiany obowiązek, precyzującego jego zakres, a także określającego sposoby jego wykonania.

W przypadku uczelni zobowiązanie do zapewniania otwartości publikacji powinno być naszym zdaniem wprowadzone przez władze uczelni w odniesieniu do wszystkich funkcjonujących w niej jednostek naukowych. Wprowadzanie go na poziomie pojedynczych jednostek organizacyjnych uczelni (choć możliwe) nie jest w naszej ocenie najbardziej efektywnym sposobem. Najlepiej, aby uczelnia wprowadziła takie zobowiązanie poprzez wyraźne ujęcie go w regulaminie zarządzania prawami autorskimi i innymi prawami na dobrach niematerialnych, do którego posiadania zobowiązuje uczelnie prawo o szkolnictwie wyższym. W przypadku instytucji naukowych innych niż uczelnie, niezobowiązanych do posiadania takiego regulaminu, może to być innego rodzaju dokument, taki jednak, którego przyjęcie skutkowałoby nałożeniem na pracowników naukowych skutecznego zobowiązania do zapewniania otwartości.

W niniejszym rozdziale wyjaśniamy, w jaki sposób wprowadzić takie zobowiązanie. W tym celu omawiamy kolejno: (1) rodzaje otwartego dostępu, jakie mogą być ujmowane w politykach otwartości, (2) relacje pomiędzy polityką otwartości a prawami autorskimi (osobistymi i majątkowymi), (3) relacje pomiędzy praktyką wydawniczą a kształtem polityki otwartości, (4) miejsce polityki otwartości w wewnętrznych regulacjach prawnych uczelni.

### 3.1. Rodzaje otwartego dostępu

Polityka otwartości może wymagać zapewnienia otwartego dostępu gratis lub libre albo pozostawiać zobowiązanym osobom wybór co do rodzaju otwartego dostępu. Otwarty dostęp gratis do publikacji naukowych to udostępnianie ich w publicznym Internecie, bez blokad technicznych i opłat. Do zapewnienia tej postaci otwartego dostępu nie jest konieczne udzielenie odbiorcom publikacji żadnej licencji; mogą oni wtedy korzystać z publikacji tylko zgodnie z właściwymi przepisami o dozwolonym użytku lub o innych wyjątkach przewidzianych w przepisach prawa. Natomiast otwarty dostęp libre zachodzi, jeżeli poza udostępnieniem publikacji w powyższy sposób zostanie do niej udzielona licencja na nieograniczone, nieodpłatne i niewyłączne korzystanie z oryginałów oraz z ich ewentualnych opracowań (wolna licencja).

Wolna licencja to taka licencja, która nie ogranicza powyższych szerokich swobód licencjobiorcy w sposób istotny. W ramach różnych funkcjonujących w praktyce definicji wolności i otwartości (Free Software Definition<sup>21</sup>, Definition of Free Cultural Works<sup>22</sup>, Budapest Open Access Initiative<sup>23</sup>) dopuszcza się tylko ograniczenia polegające na obowiązku podawania autorstwa i źródła (ew. także innych informacji o twórcy, producencie lub wydawcy, przedmiocie licencji oraz o jej postanowieniach), a niekiedy także polegające na obowiązku udostępnienia odbiorcom przedmiotu licencji lub jego opracowania na takiej samej licencji (tzw. klauzule *copyleft/share-alike*). Licencje zawierające inne ograniczenia nie są wolnymi licencjami, wobec czego udostępnione na nich publikacje nie znajdują się w otwartym dostępie libre (przy czym, w zależności od pozostałych okoliczności, może okazać się, że mamy nadal do czynienia z otwartym dostępem gratis).

### 3.2. Polityka otwartości a prawo autorskie

Udostępnienie utworu w określony sposób wymaga decyzji osoby, której przysługują prawa autorskie do danego utworu, chyba że jest dokonywane w zakresie dozwolonego użytku (przepisy o dozwolonym użytku pozwalają w pewnych przypadkach korzystać z utworów bez zgody osób uprawnionych). Nie istnieje jednak przepis o dozwolonym użytku pozwalający na swobodne udostępnianie cudzych utworów w Internecie, wobec czego o otwartym dostępie musi zdecydować uprawniony lub upoważniona przez niego do tego osoba. W związku z powyższym, uwzględnienie istnienia prawa autorskiego jest konieczne zarówno przy udostępnianiu publikacji w otwartym dostępie gratis, jak i w otwartym dostępie libre. Chodzi tu przede wszystkim o to, aby instytucja naukowa, formułując swoją politykę otwartości, nałożyła zobowiązanie do umieszczania publikacji w otwartym dostępie na osobę, której przysługują stosowne prawa autorskie, i z uwzględnieniem specyfiki tych praw. Należy tu uwzględnić fakt, że prawa autorskie dzielą się na prawa osobiste i majątkowe.

---

21 <https://www.gnu.org/philosophy/free-sw.html>

22 <http://freedomdefined.org/Definition>

23 <http://www.budapestopenaccessinitiative.org/boai-10-recommendations>

### 3.2.1. Autorskie prawa osobiste

Autorskie prawa osobiste przysługują zawsze twórcy (są niezbywalne) i chronią jego osobistą, niemajątkową więź z utworem przejawiającą się m.in. w uznaniu jego autorstwa, oznaczaniu tego autorstwa (nazwiskiem lub pseudonimem, ale też zachowania anonimowości, jeżeli taka jest wola twórcy), ochronie integralności utworu, zapewnieniu twórcy nadzoru nad sposobem korzystania z utworu, a wreszcie samodzielnym decydowaniu przez twórcę o pierwszym udostępnieniu utworu publiczności. Zobowiązanie twórcy do umieszczenia jego utworu w otwartym dostępie wkraczałoby w te prawa, gdyby powodowało naruszenie tej więzi. Byłoby tak na przykład wtedy, gdyby autorstwo twórcy zostało wbrew jego woli ukryte lub wręcz przypisane innej osobie. Inne przykłady naruszenia autorskich praw osobistych to udostępnienie utworu w postaci niezgodnej z wolą twórcy czy też zmuszanie twórcy do publikacji, zanim on sam uzna, że utwór jest do tego gotowy (np. brak autoryzacji korekty).

Wdrażanie otwartego dostępu polega na udostępnianiu w sposób otwarty prac naukowych, a nie na wymuszaniu na autorze decyzji o publikacji. W ramach polityki otwartości instytucja naukowa nie pozbawia zatem autorów prawa do decydowania o tym, co opublikują i w jakiej postaci. Wymaga natomiast, aby po podjęciu tych decyzji zapewnili otwartość swoich publikacji (niezwłocznie lub po upływie określonego terminu). To zatem nadal autor, a nie ktoś inny, decyduje o wszystkich kwestiach wpływających na jego więź z utworem.

Jeżeli chodzi o pozostałe prawa osobiste autorów, to dla ich poszanowania w otwartym dostępie kluczowe jest prowadzenie przez instytucję zgodnego ze standardami i dobrymi praktykami otwartego repozytorium. Zapewnia to poszanowanie integralności publikacji i jej powiązanie z prawidłowymi metadanymi (szczególnie z informacjami o autorstwie i afiliacji autora).

Ograniczenia ani wkroczenia w prawa osobiste nie stanowi w szczególności decyzja o udostępnieniu utworu w otwartym dostępie libre. Udzielana wtedy wolna licencja nie prowadzi do pozbawienia autora praw osobistych (są one niezbywalne), a klauzule uznania autorstwa pomagają wręcz ustalić prawidłowy sposób ich przestrzegania. Dla przykładu, standardowe wzorce wszystkich licencji Creative Commons zawierają wyraźne postanowienia zobowiązujące do przestrzegania praw osobistych. Poza doprecyzowaniem sposobu przestrzegania prawa do autorstwa wymagają one również od licencjobiorcy, aby wyraźnie zaznaczył, jeżeli udostępnia opracowanie, a nie oryginał. Nie może on też korzystać z utworu w żaden sposób, który sugerowałby poparcie lub aprobatę licencjodawcy dla jego działań w stosunku do utworu. Nawet jeżeli uznamy, że opracowywanie narusza integralność utworu, to licencje CC wyraźnie wymagają, aby czynić to z poszanowaniem więzi łączącej autora oryginału z jego dziełem. Oznacza to, że ewentualne wymaganie od twórcy, aby po podjęciu decyzji o publikacji udzielił on każdemu zainteresowanemu wolnej licencji, nie pozbawia go jego prawa do autorstwa ani też pozostałych praw osobistych, takich jak prawo do integralności. W żadnym wypadku nie zmusza go to do tolerowania naruszeń autorskich praw osobistych.

Niekiedy pojawiają się obawy, że otwarty dostęp narazi autorów na większe ryzyko plagiatów. Publiczna dostępność utworu, a zwłaszcza jego dostępność w Internecie, niewątpliwie powoduje, że w sensie faktycznym naruszenie autorskich praw osobistych twórcy jest łatwe i możliwe. Należy się zatem liczyć



z ryzykiem plagiatów, przeinaczeń itd. Ale im szerzej dostępny jest pełny tekst utworu, tym łatwiej takie naruszenia zidentyfikować i eliminować. Co więcej, udzielenie wolnej licencji pozwala bez prawnych wątpliwości śledzić plagiaty z użyciem technik maszynowych, co znacząco ułatwia ich wykrywanie. Otwarty dostęp ułatwia zatem wykrycie i udowodnienie naruszeń, takich jak plagiat, zwłaszcza jeżeli udostępniane są pełne teksty i odbywa się to w sposób umożliwiający mechaniczne indeksowanie i przeszukiwanie. Zatem jeżeli nawet otwarty dostęp (a szerzej: Internet) zwiększa skalę potencjalnych i faktycznych naruszeń (co nie jest jednak oczywiste, gdyż potencjalny naruszciciel widząc, że tekst jest w otwartym dostępie, może od planowanego naruszenia odstąpić z uwagi na mniejsze szanse powodzenia), to jednocześnie zwiększa się skala dostępnych metod przeciwdziałania im<sup>24</sup>.

### 3.2.2. Autorskie prawa majątkowe

Uwzględnienie autorskich praw majątkowych w realizacji polityki otwartości wymaga osobnego wytłumaczenia. Prawa te różnią się od autorskich praw osobistych pod wieloma względami – przede wszystkim tym, że są zbywalne. Zobowiązanie do udostępnienia utworu w otwartym dostępie niewątpliwie wchodzi w relację z tymi prawami, gdyż wykonanie tego zobowiązania polega na odpowiednim zadysponowaniu nimi. W przypadku otwartego dostępu gratis dysponent praw majątkowych godzi się na umieszczenie utworu w Internecie, a w przypadku otwartego dostępu libre udziela ponadto wolnej licencji. Zobowiązanie powinno zatem określać, czy chodzi o otwarty dostęp gratis, czy libre (w tym drugim wypadku należy także uwzględnić informację o licencji), a ewentualne pozostawienie tu autorom swobody wyboru powinno być dobrze przemyślane.

Z uwagi na zbywalność autorskich praw majątkowych istotne jest w pierwszej kolejności, aby omawiane zobowiązanie zostało nałożone na właściwą osobę. Autorskie prawa majątkowe powstają co do zasady na rzecz twórcy, chyba że co innego wynika wyraźnie z ustawy. Te ustawowe wyjątki będą jednak dotyczyły publikacji naukowych tylko w bardzo specyficznych przypadkach (np. publikacja kodu źródłowego bądź wynikowego programu komputerowego stworzonego przez pracownika). Warto tu wyraźnie przypomnieć, że prawa te powstają po stronie twórcy, nawet jeżeli publikacja naukowa została stworzona w ramach jego zatrudnienia w instytucji naukowej (ustawa daje wtedy tej instytucji jedynie pierwszeństwo publikacji).

### 3.3. Polityka otwartości a praktyki wydawnicze

Ponieważ autorskie prawa majątkowe są zbywalne, to mogą zostać przeniesione na inne podmioty (np. na wydawcę), przez co dla trwałego umieszczenia publikacji w otwartym dostępie nie wystarcza samo tylko zobowiązanie twórcy jako pierwotnego podmiotu praw. Polityka otwartości powinna uwzględniać fakt przenoszenia praw majątkowych przez twórców na inne podmioty. Z punktu widzenia autorów, na rzecz których powinna działać polityka otwartości, optymalnym długofalowym rozwiązaniem jest sytuacja, w której autorzy zachowują swoje prawa, udzielając wydawcom jedynie licencji. Jednak w krótkim okresie polityka otwartości nie powinna całkowicie zamykać autorom drogi do publikowania w wydawnictwach, które wymagają przenoszenia praw.

---

24 <http://www.plagiarismadvice.org/research-papers/item/plagiarism-detection-in-open-access-publication>

Jest niezwykle istotne, aby pomimo przeniesienia praw na takie wydawnictwo zachowana została możliwość umieszczenia publikacji w otwartym dostępie. Przy obecnym brzmieniu ustawy o prawie autorskim nie jest bowiem jasne, czy po przeniesieniu praw udzielone wcześniej licencje wygasają, czy nie. Instytucja naukowa nie może zatem poprzestać na pozyskaniu licencji na otwarty dostęp i nie interesować się dalszym przenoszeniem praw przez autora. W celu uniknięcia wątpliwości co do zachowania takiej licencji, warto aby nowy uprawniony wyraźnie uznał jej skuteczność. Wymaga to co najmniej, aby został do tego wyraźnie zobligowany w umowie z dotychczasowym uprawnionym.

Istotnym ułatwieniem jest w tym przypadku fakt, że wiele wydawnictw naukowych uwzględnia otwarty dostęp w swojej działalności, nawet jeżeli jednocześnie wymaga od autorów przenoszenia praw. Dość powszechną praktyką takich wydawnictw jest obecnie udzielanie autorom wyraźnych zezwoleń na udostępnienie publikacji w repozytorium w ramach tzw. zielonej drogi (może to być ograniczone do wersji wcześniejszej niż ostateczna wersja opublikowana lub objęte określonym terminem karencji – jest to tzw. embargo). Niektórzy wydawcy wykorzystują także tzw. złotą drogę, samemu umieszczając wydawane przez siebie publikacje w otwartym dostępie. Nie musi się to automatycznie przekładać na możliwość udostępnienia publikacji w repozytorium instytucjonalnym, ale może pozwolić na osiągnięcie celu polityki otwartości. Polityka otwartości powinna zatem stymulować autorów do korzystania z czasopism takich wydawców, którzy umożliwiają otwarty dostęp na drodze zielonej lub złotej.

Sytuacja jest najmniej kłopotliwa w przypadku wydawców, którzy nie wymagają przenoszenia praw i poprzestają na uzyskaniu od autora jedynie niewyłącznej licencji – autor może wtedy co do zasady swobodnie wywiązać się ze zobowiązania do zapewnienia otwartego dostępu (rozstrzygają jednak konkretne postanowienia zawartej umowy; nadal też pozostaje ryzyko późniejszego przeniesienia praw przez autora).

Poza wydawcami wykorzystującymi zieloną lub złotą drogę otwartego dostępu występują też oczywiście wydawcy, którzy wymagają przenoszenia praw i jednocześnie nie pozwalają na żadną formę otwartego dostępu ani sami go nie realizują. Polityka otwartości nie powinna z jednej strony całkowicie wyłączać możliwości publikacji w takich wydawnictwach, zwłaszcza jeżeli nie istnieje akurat wydawnictwo otwarte o równie wysokiej jakości. Z drugiej jednak strony w ramach tej polityki można przewidzieć różne rozwiązania tego problemu. Przede wszystkim, jak pokazują światowe doświadczenia, już sama instytucjonalna polityka otwartości stanowi istotny czynnik powodujący, że wydawcy zmieniają lub dostosowują swoje wzory umów i praktyki. W sytuacjach gdy – pomimo informacji o związaniu autora polityką – wydawca nie reaguje, instytucja naukowa powinna zasugerować autorowi możliwość publikacji w otwartych, a jednocześnie zapewniających wysoką jakość wydawnictwach. Udostępnieniu autorom informacji o takich wydawnictwach (baza SHERPA/ROMEO to źródło podstawowej wiedzy na ten temat) powinno towarzyszyć wymaganie, aby autor, chcąc skorzystać z wydawnictwa zamkniętego, musiał to co najmniej wyczerpująco uzasadnić, a może wręcz uzyskać wyraźną zgodę przełożonego. Instytucja naukowa powinna ponadto przygotować dla swoich pracowników gotowe wzory umów wydawniczych lub aneksów do już zawartych umów wydawniczych, tak aby wspomóc ich w ten sposób w negocjacjach z wydawnictwami, które same nie realizują otwartego dostępu. Umowy te nie ograniczałyby wydawnictw w ich działalności (np. wydawanie

drukiem, publikacje elektroniczne), a twórcy w udostępnianiu utworu za pomocą repozytorium (wyłącznie do dozwolonego użytku – w modelu gratis – lub wraz z udzieleniem wolnej licencji – w modelu libre).

Niezależnie od stanowiska danego wydawcy odnośnie do otwartego dostępu, z uwagi na praktykę przenoszenia praw na wydawców zobowiązanie autorów do przenoszenia autorskich praw majątkowych na instytucję naukową nie jest najlepszym rozwiązaniem, jeżeli miałyby objąć publikacje przeznaczone do publikacji w zewnętrznym wydawnictwie, z którym ta instytucja nie współpracuje. Instytucja musiałaby się wtedy zaangażować w rozmowy z takim wydawnictwem, co nie wydaje się praktyczne, ale może być warte rozważenia, jeśli dzięki temu mogłoby dojść do wynegocjowania korzystniejszych warunków dla autorów i jednostki naukowej. Niewątpliwym plusem takiego rozwiązania jest fakt, że instytucja naukowa mogłaby wtedy sama decydować o umieszczaniu publikacji w otwartym dostępie w sposób sprawny i trwały (po przeniesieniu praw autor nie ma łatwych możliwości doprowadzenia do ich powrotu).

Przeniesienie praw na instytucję naukową może być natomiast zalecane w sytuacji, w której publikacja naukowa jest realizowana wysiłkiem lub z udziałem samej instytucji naukowej. Będzie tak na przykład wtedy, gdy instytucja prowadzi projekt badawczy, w ramach którego wchodzi we współpracę z wydawnictwem lub wręcz sama pełni rolę wydawcy i w związku z tym odpowiada za pozyskiwanie praw. Jednak nawet wtedy przeniesienie praw na instytucję lub wydawcę nie jest niezbędne i istnieje rozwiązanie, w ramach którego prawa te pozostają przy twórcy, a jednocześnie publikacja zostaje wydana oraz umieszczona w otwartym dostępie. Rozwiązanie to polega na wprowadzeniu odpowiednio sformułowanej licencji oraz zobowiązania do jej udzielenia.

### **3.4. Polityka otwartości w hierarchii wewnętrznych aktów prawnych na uczelni**

Niezależnie od przyjętych rozwiązań, powinny one wraz z całą polityką otwartości wpisywać się w już istniejącą relację prawną pomiędzy autorem publikacji a instytucją naukową. Relacja ta powinna istnieć, zanim dojdzie do powstania konkretnej pracy naukowej, natomiast jest konieczna przed podjęciem kroków mających na celu jej opublikowanie. Relacja jest stosunkiem prawnym, w wyniku którego dochodzi do przygotowania publikacji (mowa przecież o publikacjach powstających w ramach zatrudnienia w danej instytucji naukowej, a nie w wyniku osobnych aktywności naukowców). W przypadku pracowników będzie to stosunek pracy, a w przypadku pozostałych osób najprawdopodobniej umowa o dzieło (nie można też wykluczyć sytuacji, w której pracownik zawiera ze swoim pracodawcą osobną umowę o dzieło na przygotowanie publikacji niemieszczącej się w jego dotychczasowych obowiązkach pracowniczych). Zgodnie z art. 86c ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (Dz. U. Nr 164, poz. 1365 z późn. zm.) zasady dotyczące obu tych rodzajów umów na uczelni powinien porządkować regulamin wydany na podstawie tego przepisu.

Z regulaminu (choć w przypadku instytucji naukowych innych niż uczelnie nie będzie to regulamin w rozumieniu art. 86c ww. ustawy) powinno wynikać wprost, że twórcy publikacji naukowych powstających w wyniku umów zawieranych na podstawie tegoż regulaminu są zobowiązani do umieszczania ich w otwartym dostępie. Dzięki takiemu rozwiązaniu polityka otwartości zostanie wprowadzona w sposób względnie trwały i jednolity. Takie

zobowiązanie może być ewentualnie wprowadzone na niższym poziomie (np. w postaci zarządzenia kierownika jednostki naukowej uczelni zawierającego wytyczne co do treści umów czy – w ostateczności – indywidualnie w treści każdej pojedynczej umowy czy aktu mianowania), jednak regulamin nie może wtedy tego wykluczać. Takie rozwiązanie niesie za sobą wszakże ryzyko bałaganu organizacyjnego.

Rozwiązaniem docelowym jest wobec tego takie sformułowanie postanowień regulaminu dla całej instytucji naukowej, aby autorzy publikacji powstających w ramach tej instytucji byli zobowiązani do umieszczania ich w otwartym dostępie gratis lub libre w repozytorium instytucjonalnym niezależnie od tego, jakie umowy dotyczące publikacji autor zawarze z osobami trzecimi (chodzi o udostępnienie w repozytorium pełnego tekstu). Regulamin na zasadzie wyjątku może dopuszczać odstępstwa od tego zobowiązania. Natomiast w stosunku do niektórych publikacji można rozważyć osobne rozwiązanie, zgodnie z którym autorzy byłiby zobowiązani do przeniesienia praw na instytucję – dotyczy to już wspomnianych publikacji wydawanych przez samą instytucję naukową (np. uczelniane wydawnictwo) lub we współpracy z wydawnictwem zewnętrznym. Otwarty dostęp będzie wtedy realizowany przez samą instytucję naukową lub wydawnictwo zgodnie z przyjętymi uzgodnieniami.

Postanowienia regulaminu powinny zatem wyraźnie regulować następujące kwestie: po pierwsze, powinny zawierać zobowiązanie do deponowania publikacji naukowej w instytucjonalnym repozytorium niezwłocznie po przyjęciu jej przez wydawnictwo lub też po wydaniu, wraz z jednoczesnym zapewnieniem otwartego dostępu (ewentualnie otwarcie zdeponowanej publikacji po upływie odpowiedniego terminu, czyli embarga). Zobowiązanie to obejmować powinno ponadto udzielenie odpowiedniej licencji instytucji naukowej, najlepiej z jednoczesnym określeniem jej postanowień oraz wskazaniem momentu, w którym zostaje ona udzielona. Osobno należy uwzględnić ewentualne zobowiązanie do zapewnienia otwartego dostępu libre odbiorcom publikacji, najlepiej poprzez wskazanie konkretnych standardowych licencji do wyboru, takich jak wolne licencje Creative Commons. Istotne jest tu także zobowiązanie autora, aby powstrzymał się od przenoszenia praw w sposób, który mógłby spowodować, że wzmiankowane licencje utracą swoją moc obowiązującą. Autor powinien być zatem co najmniej zachęcany do skorzystania z oferowanych przez wydawców opcji otwartego dostępu na drodze złotej lub zielonej, a jeżeli wydawca takiej opcji nie oferuje – do podjęcia negocjacji umowy ze wsparciem instytucji naukowej.

Po drugie, powinno się w nich znaleźć ewentualne zobowiązanie do przeniesienia na instytucję praw, również w odpowiednio określonym zakresie, w przypadku specjalnej kategorii publikacji (np. wzmiankowane wyżej projekty badawcze, w ramach których kwestię publikacji organizuje sama instytucja).

Konkretne sformułowania użyte w regulaminie mogą się różnić w zależności od tego, czy są skierowane do pracowników, czy też do osób zatrudnionych na innej podstawie (np. publikacje powstające w ramach umów o dzieło). Prawo autorskie różnicuje bowiem sytuację tych dwóch grup: w odniesieniu do pracowniczych utworów naukowych art. 14 pr. aut. pierwszeństwo publikacji przyznaje instytucji naukowej zatrudniającej pracownika. Z kolei w odniesieniu do utworów naukowych powstałych w wyniku umów innych niż umowy o pracę instytucja nie ma takiego pierwszeństwa.

Zastosowanie tego przepisu wprost polegałoby na egzekwowaniu pierwszeństwa publikacji przysługującego instytucji naukowej, które realizowane byłoby poprzez udostępnienie w Internecie. W takim przypadku twórcy musieliby otrzymywać wynagrodzenie za skorzystanie przez instytucję naukową z prawa pierwszeństwa (obowiązek zapłaty wynagrodzenia wynika wyraźnie z art. 14 pr. aut.). Ponadto zastosowanie tego przepisu wprost w celu wprowadzenia otwartego mandatu byłoby obciążone pewnym ryzykiem z uwagi na wątpliwości związane z użytym w nim sformułowaniem „publikacja”. Art. 6 ust. 1 pkt 1 pr. aut. definiuje bowiem utwór opublikowany poprzez pojęcie egzemplarza, którym jest fizycznie istniejący nośnik. Pojawia się wobec tego wątpliwość, czy „publikacja” obejmuje udostępnianie w Internecie, czyli w sytuacji, gdy nie dochodzi do powstawania takich nośników, przynajmniej na pozór (technicznie rzecz biorąc, ma miejsce kopiowanie z komputera na komputer). Przy ostrożnej interpretacji pierwszeństwo publikacji obejmowałoby jedynie media tradycyjne (np. druk).

Istnieją także wątpliwości, czy stosowanie art. 14 pr. aut. wprost pozwoli na wprowadzenie modelu libre, gdyż art. 14 ust. 1 pr. aut. nie daje instytucji naukowej prawa do licencjonowania pracowniczych utworów naukowych, a jedynie pierwszeństwo ich publikacji. Wreszcie stosowanie art. 14 pr. aut. wprost może zamknąć autorom drogę do publikacji ich utworów w czasopiśmie, które nie przyjmują do wydania prac już wcześniej publikowanych (co w praktyce jest równoznaczne z niemożliwością publikacji w wielu czasopiśmie wysokiej jakości).

Omawiany przepis stanowi jednak wyraźnie, że umowa z autorem może regulować objęte nim kwestie odmiennie. Instytucja naukowa nie musi ponadto korzystać z przysługującego jej pierwszeństwa. Może ona pozwolić pracownikowi na publikację za pomocą swobodnie wybranej osoby trzeciej (np. wydawnictwa). Takie zezwolenie stanowi świadczenie instytucji naukowej na rzecz pracownika (rezygnację z pierwszeństwa), w zamian za które to ona, a nie twórca może oczekiwać określonego świadczenia wzajemnego. Jako świadczenie wzajemne za zezwolenie na publikację za pomocą swobodnie wybranego wydawcy pracownik może zostać zobowiązany do tego, aby umieścić utwór w prowadzonym przez instytucję naukową repozytorium służącym do udostępniania utworów naukowych w modelu gratis lub libre. Uważamy, że niekiedy wprost formułowane oczekiwanie, aby pracownicy naukowcy publikowali w jak najwyższej punktowanych czasopiśmie, nie wpływa na prezentowane tu rozwiązanie.

Powyższa rezygnacja z pierwszeństwa nie stoi oczywiście na przeszkodzie jednoczesnemu przyjęciu przez wewnętrzne (np. uczelniane) wydawnictwo zasady, że publikacje udostępniane są w otwartym dostępie. Taka zasada powinna znaleźć odzwierciedlenie we wzorach umów stosowanych w ramach takiego wydawnictwa.

Dostosowaniu regulaminu oraz uzupełnieniu wzorów umów musi towarzyszyć umożliwienie realizacji polityki otwartości w sensie organizacyjno-technicznym. W tym celu instytucja naukowa powinna prowadzić repozytorium, w którym będzie można umieszczać publikacje autorów afiliowanych przy tej instytucji (więcej o prowadzeniu repozytorium od strony technicznej piszemy w rozdziale piątym niniejszego raportu).

# Rozdział 4

## Działalność wydawnicza

W niniejszym rozdziale przedstawiamy rekomendacje dotyczące działalności wydawniczej, skierowane do instytucji naukowych zainteresowanych wdrożeniem otwartych modeli dystrybucji treści. Zaczynamy od propozycji określenia podstawowych zasad, na których może opierać się owo wdrożenie, a następnie przedstawiamy bardziej szczegółowe zalecenia dotyczące stosunków prawnych między wydawcą a autorami oraz technicznych aspektów dystrybucji treści naukowych w Internecie.

Prezentowane rekomendacje mają stanowić pomoc dla tych instytucji, które są zdecydowane wykorzystać szanse oferowane przez otwarty dostęp. Pragniemy podkreślić, że nie dotyczą one samego procesu wydawniczego, zapewnienia wysokiej jakości publikacji – zarówno pod względem merytorycznym, jak i redakcyjnym – czy innych kwestii niezwiązanych bezpośrednio z dystrybucją treści.

Zasady polityki wydawniczej przyjęte przez daną instytucję powinny być czytelne i przedstawione autorom w zrozumiałym sposób. Autor powinien zostać w szczególności poinformowany o polityce wydawcy w stosunku do samoarchiwizacji publikacji w repozytorium: którą wersję tekstu może zdeponować (np. ostateczną wersję autorską, ostateczną wersję opublikowaną) i w jakim terminie (od razu po przyjęciu tekstu, w momencie publikacji, dopiero po okresie embargo).

### 4.1. Cele działalności wydawniczej

Przed podjęciem decyzji o wprowadzeniu zmian w obecnym modelu dystrybucji publikacji naukowych warto przemyśleć ostateczne cele działalności wydawniczej instytucji naukowej. Pozwoli to uniknąć sytuacji, w której cele drugorzędne, lecz możliwe do realizacji w krótkiej perspektywie, przesłonią cele podstawowe, osiągalne tylko długofalowo. Wydaje się, że właściwa priorytetyzacja celów znalazła odzwierciedlenie np. w zarządzeniu rektora UMK ws. zasad finansowania działalności wydawniczej oraz cyfrowej Platformy Czasopism Uniwersytetu Mikołaja Kopernika w Toruniu, gdzie mowa jest o tym, że Uniwersytet prowadzi działalność wydawniczą „[w] celu wspierania działalności naukowej oraz kształcenia kadr (...)”<sup>25</sup>.

<sup>25</sup> Por.: [https://www.umk.pl/uczelnia/dokumenty/biuletyn/prawo/?akcja=dokument&typ=Z\\_Rektora&nr=1&bp=1&rok=2013](https://www.umk.pl/uczelnia/dokumenty/biuletyn/prawo/?akcja=dokument&typ=Z_Rektora&nr=1&bp=1&rok=2013). Inaczej priorytety wydawnictwa uczelnianego określa np. zarządzenie rektora Uniwersytetu Warszawskiego: „Podstawowym zadaniem Wydawnictw jest opracowywanie, publikacja i sprzedaż książek

Wśród naczelnych zasad wyznaczających kierunek rozwoju instytucji naukowej w zakresie działalności wydawniczej powinno znajdować się wsparcie zarówno samej instytucji, jak i jej pracowników, przy jednoczesnej odpowiedniej hierarchizacji celów.

#### **4.1.1. Wspieranie pracowników naukowych w budowaniu pozycji, prestiżu i spełnianiu wymagań instytucji finansujących badania**

Publikacje wydawane przez instytucję naukową powinny być dystrybuowane w taki sposób, aby zmaksymalizować szansę dotarcia do czytelników, zwłaszcza naukowców z tej samej co autor dyscypliny, ponieważ jest to warunek osiągnięcia przez autora rezultatów oczekiwanych w czasie pracy nad publikacją – zarówno wyższych wskaźników cytowań, jak i prestiżu.

W badaniu polskich naukowców przeprowadzonym na potrzeby raportu *Otwarta Nauka w Polsce 2014. Diagnoza* uwzględniona została kwestia kryteriów wyboru miejsca publikacji przez badaczy<sup>26</sup>. Bardzo ważne okazały się m.in. współczynnik wpływu, punktacja i prestiż periodyku. Prawie 60% respondentów wskazało na dostępność publikacji w Internecie, co świadczy o wadze, jaką autorzy przywiązują do szerokiego rozpowszechniania swoich publikacji. Raport STM (International Association of Scientific, Technical and Medical Publishers) z 2015 roku, podsumowując badania dotyczące różnorodnych aspektów komunikacji naukowej, informuje, że główną motywacją naukowców przy podjęciu decyzji o publikacji jest chęć upowszechnienia wyników badań, natomiast wśród najważniejszych motywacji drugorzędnych znajdują się chęć rozwoju kariery i zwiększenia szans na przyszłe finansowanie<sup>27</sup>.

Z punktu widzenia autora wszystko, co utrudnia mu dotarcie do czytelników, jest przeszkodą w osiągnięciu celu, który przyświecał mu podczas pracy nad publikacją. Szeroki dostęp do publikacji przynosi korzyść zarówno jeżeli chodzi o zwiększenie liczby cytowań, jak i w mniej precyzyjnie określonym wymiarze prestiżu. Prestiż wymaga bowiem znajomości publikacji przez środowisko naukowe<sup>28</sup>.

Otwarty dostęp do publikacji naukowych może być z jednej strony preferencją samych autorów, z drugiej zaś – obowiązującym ich wymogiem. Instytucje finansujące badania coraz częściej zobowiązują naukowców do zapewnienia otwartego dostępu do publikacji powstałych w wyniku finansowanych przez nie projektów – takie zasady przyjęła np. Komisja Europejska w programie Horyzont 2020. Działalność wydawnicza powinna być więc prowadzona z uwzględnieniem zarówno potrzeb autorów, jak i ciążących na nich obowiązków.

#### **4.1.2. Wspieranie instytucji naukowej we wzmacnianiu jej pozycji w środowisku akademickim i poza nim**

Szerokie udostępnianie publikacji leży w interesie nie tylko autorów, lecz również instytucji, których są oni pracownikami. Działalność wydawnicza instytucji może stać się instrumentem budowania, wzmacniania

---

oraz czasopism (...)" . Por.: <http://monitor.uw.edu.pl/Lists/Uchway/Attachments/411/M.2012.104.Zarz.20.pdf>.

26 *Otwarta Nauka w Polsce 2014. Diagnoza*, red. J. Szprot, Warszawa 2014, s. 83–86.

27 *The STM Report*, 4th edition, s. 70–71.

28 Por.: A. Mas-Bleda, M. Thelwall, K. Kousha, I.F. Aguillo, *Do highly cited researchers successfully use the social web?*, „Scientometrics” (2014), 101. Dostępny pod adresem: <http://www.scit.wlv.ac.uk/~cm1993/papers/DoHCRSuccessfullySocialWeb.pdf>. Spośród ponad 1500 naukowców z Europy o najwyższym wskaźniku cytowań 970 (64%) miało własną stronę internetową, a 28% było aktywnych w akademickich mediach społecznościowych (m.in. Mendeley, Academia.edu).

i utrzymywania jej rozpoznawalności oraz renomy, a co za tym idzie przyczynić się do zwiększenia możliwości współpracy z innymi instytucjami, również spoza obszaru nauki (np. przedsiębiorstwami). W kwestii dystrybucji publikacji interes instytucji naukowej jest więc zbieżny z interesem indywidualnego naukowca (oraz zespołów badawczych). Instytucja we własnym interesie powinna zatem usuwać przeszkody w rozpowszechnianiu swoich publikacji.

Różnice między interesem autora i instytucji mogą pojawić się przede wszystkim wtedy, gdy instytucji zależy na zbudowaniu prestiżowego czasopisma, co wymaga długotrwałych starań m.in. o zwiększanie liczby cytowań. Naukowiec może nie chcieć poświęcać wyników własnej pracy na takie działania i zdecydować się na publikację w czasopiśmie mającym już silną pozycję. Ten problem nie dotyczy jednak wyłącznie czasopism wydawanych w modelu otwartym, ale każdego nowego periodyku.

#### **4.1.3. Hierarchizacja celów i korzyści**

Instytucja naukowa w swojej działalności wydawniczej powinna określić cele podstawowe oraz cele drugorzędne. W szczególności zachęcamy do przemyślenia relacji między zwiększaniem dostępności, widoczności i liczby cytowań publikacji a dążeniem do uzyskania przychodu ze sprzedaży publikacji. Przychód ze sprzedaży w przypadku płatnego dostępu pokrywa na ogół niewielką część kosztów, nie sprawia więc, że publikacja uzyskuje sens ekonomiczny. Płatny dostęp, niezależnie od wysokości opłaty, ogranicza natomiast zasięg publikacji. W konsekwencji zdarza się, że dążenie do uzyskania – mało znaczącego w porównaniu z poniesionymi kosztami – przychodu wyraźnie utrudnia realizację podstawowego celu publikowania, czyli dotarcia do szerokiego grona badaczy oraz polepszenia pozycji naukowca i instytucji w środowisku akademickim. W cytowanym wcześniej zarządzeniu rektora UMK ewentualne przychody nie są określone jako cel działalności wydawniczej i wydaje się, że jest to świadoma decyzja.

## **4.2. Kwestie prawnoautorskie**

W działalności wydawniczej ważną rolę odgrywają kwestie prawnoautorskie. Wydawca może opublikować tekst wyłącznie wtedy, gdy uzyska stosowne uprawnienia. Zdobycie ich możliwe jest jednak na różne sposoby. Wydawca powinien wybrać taki, który przyniesie korzyści zarówno jemu, jak i autorom.

Może się zdarzyć, że instytucja naukowa wprowadzi zobowiązanie do udostępnienia publikacji w sposób otwarty (o takim rozwiązaniu piszemy w rozdziale trzecim). Jednostka odpowiedzialna za działalność wydawniczą instytucji powinna wypracować swoje stanowisko w zakresie otwartego dostępu niezależnie od takiego zobowiązania (oczywiście jeśli takie zobowiązanie istnieje, to przyjęte stanowisko powinno być z nim zgodne), ponieważ zazwyczaj publikuje ona także prace spoza swojej macierzystej instytucji.

Na oczekiwania autorów wobec wydawców w kwestiach prawnoautorskich wpływ mogą mieć otwarte mandaty wprowadzane przez instytucje wobec pracowników lub otwarte mandaty wprowadzane wobec beneficjentów przez instytucje finansujące badania.



#### 4.2.1. Stosunki prawnoutorskie między wydawcą a autorami

Relacje prawnoutorskie między autorem a wydawcą powinny zostać ustalone w taki sposób, by wspierały interes obu stron. Ponieważ jak najszersze upowszechnianie publikacji leży w interesie zarówno wydawcy, jak i naukowca, relacje prawnoutorskie powinny umożliwiać im realizację tego celu. Przede wszystkim zaś umowa zawierana przez wydawcę i autora nie powinna ograniczać autorowi – w stopniu większym niż wymaga tego interes wydawcy – możliwości korzystania z narzędzi ułatwiających upowszechnianie tekstu.

Można to osiągnąć na kilka sposobów: poprzez przeniesienie praw na wydawcę z udzieleniem autorowi licencji zwrotnej lub poprzez udzielenie przez autora licencji z prawem sublicencjonowania. Wydawca powinien określić szczegółowe cele, które chce osiągnąć poprzez zawarcie umowy z autorem. Cele te powinny uwzględniać obecnie istniejące pola eksploatacji utworu (postanowienia dotyczące nieznanych dziś pól eksploatacji będą nieskuteczne). Wydawca powinien uwzględnić, że w przyszłości może zdecydować o wykorzystaniu utworu w sposób inny niż obecnie (na polu eksploatacji już istniejącym w momencie podpisywania umowy z autorem). Z drugiej strony wydawca powinien rozważyć stopień komplikacji procesu zawierania umów, tzn. ocenić, czy woli zawierać umowy w formie pisemnej, czy raczej z wykorzystaniem np. elektronicznego formularza.

Zależnie od zamierzonych celów, wydawca może proponować autorom albo umowę przenoszącą majątkowe prawa autorskie do utworu na wydawcę (taka umowa wymaga formy pisemnej), albo licencję, która uprawnia wydawcę do publikacji tekstu, ale pozostawia prawa przy autorze<sup>29</sup>. Coraz większą popularnością wśród wydawców naukowych cieszą się licencje Creative Commons<sup>30</sup>, które pozwalają udzielić standardowej licencji według jednego z sześciu wzorów różniących się zakresem uprawnień przyznawanych użytkownikom/czytelnikom. Stosowanie standardowych licencji upraszcza proces obsługi prawnej i zmniejsza jego koszty, ponieważ usuwa konieczność formułowania własnej licencji (z punktu widzenia licencjodawcy) oraz zapoznawania się z postanowieniami danej licencji i porównywania ich z innymi (z punktu widzenia licencjobiorcy). Dwie spośród licencji Creative Commons określane są mianem „wolnych licencji”: CC BY (Uznanie autorstwa) oraz CC BY-SA (Uznanie autorstwa–Na tych samych warunkach), ponieważ zapewniają odbiorcom najszerszy zakres uprawnień. Odradzamy stosowanie własnych, niestandardowych wolnych licencji zastępujących licencje Creative Commons.

Bez względu na przyjęte rozwiązania rekomendujemy, aby wydawcy uwzględniali możliwość udostępnienia opublikowanego utworu naukowego w otwartym repozytorium przez jego autora. W uzasadnionych przypadkach wydawca może rozważyć wprowadzenie embargo, polegającego na zgodzie na udostępnieniu tekstu przez autora dopiero po upływie określonego czasu. Przyjęte rozwiązania powinny być zgodne z okresami embargo dopuszczanymi przez europejskie i światowe instytucje naukowe i finansujące badania, które przyjęły otwarte mandaty (np. Komisję Europejską). Przyjęcie standardowego okresu embargo umożliwia autorom wywiązanie się ze zobowiązań nakładanych na nich przez instytucje finansujące i uczelnie, zmniejsza koszty zarządzania informacją oraz sprawia, że czytelnicy wiedzą, kiedy mogą spodziewać się dostępności danego tekstu.

29 Por. K. Siewicz, *Otwarty dostęp do publikacji naukowych. Kwestie prawne*, Warszawa 2012, s. 53–59.

30 <http://creativecommons.pl/poznaj-licencje-creative-commons/>

W odniesieniu do artykułów publikowanych w czasopiśmie zazwyczaj dopuszczane jest embargo nie dłuższe niż 6 miesięcy od daty publikacji, a w naukach humanistycznych i społecznych – 12 miesięcy. Rekomendujemy umożliwianie autorom deponowania ostatecznej wersji autorskiej artykułu w repozytorium w momencie pozytywnej decyzji o publikacji (po recenzji), a ostatecznej wersji opublikowanej – w momencie publikacji.

Wyżej opisane zasady powinny odnosić się zarówno do autorów zatrudnionych w danej instytucji, jak i do autorów zewnętrznych.

### 4.3. Kwestie finansowe

Wprowadzenie otwartych modeli dystrybucji publikacji naukowych w wielu przypadkach będzie się wiązało ze zmianami w zakresie finansowania działalności wydawniczej. Ze względu na zróżnicowane rozwiązania stosowane w tym zakresie w polskich instytucjach naukowych nie sposób przedstawić jednolitych i wyczerpujących rekomendacji dotyczących tej kwestii. Z jednej strony otwarte modele pozwalają zredukować pewne koszty (np. druk, dystrybucja i przechowywanie wersji papierowych). Z drugiej strony wprowadzenie zmian może wiązać się z koniecznością poniesienia nowego rodzaju wydatków, zwłaszcza w początkowym okresie (np. związanych z wdrożeniem odpowiednich rozwiązań informatycznych). Każda instytucja powinna starannie przeanalizować te kwestie w odniesieniu do własnego modelu działania i finansowania.

#### 4.3.1. Opłaty za publikacje

Publikowanie artykułów w otwartych czasopiśmie często wiąże się z opłatą, tzw. APC (*article processing charge*). Wysokość tej opłaty jest bardzo zróżnicowana (według jednego z badań wynosi od 8 USD do 3900 USD za artykuł, przy średniej opłacie 904 USD<sup>31</sup>). W Polsce połowa czasopiśm udostępnia bezpłatnie numery bieżące<sup>32</sup>; większość z nich nie pobiera opłat od autorów. Oznacza to, że istniejący w Polsce model finansowania czasopiśm naukowych pozwala wydawać periodyki – również w modelu otwartym – bez ponoszenia dodatkowych kosztów ze strony autora. Nie można jednak wykluczyć sytuacji, w której wydawanie czasopiśma będzie wymagało wykorzystania tego finansowania. Przy zastosowaniu opłat za publikację artykułu kluczowe jest zapewnienie dostępu na wolnej licencji, aby w pełni wykorzystać możliwości stwarzane przez otwarty dostęp libre.

### 4.4. Kwestie techniczne związane z udostępnianiem publikacji elektronicznych

Zagadnienia związane z techniczną stroną udostępniania publikacji elektronicznych są równie istotne, co kwestie prawne. O ile uregulowanie stosunków prawnych umożliwia legalne działania wydawcy i autorów, o tyle zastosowanie określonych technicznych rozwiązań dotyczących dystrybucji decyduje o tym, czy publikacje są widoczne w sieci i czy optymalnie wykorzystuje się możliwości, jakie daje otwarty dostęp.

31 D.J. Solomon, B.-C. Björk, (2012), *A study of open access journals using article processing charges*, „J. Am. Soc. Inf. Sci.”, nr 63: s. 1485–1495. doi: 10.1002/asi.22673.

32 Por. *Otwarta Nauka w Polsce 2014. Diagnoza*, op. cit., s. 57.

Istnieją sprawdzone, bezpłatne rozwiązania informatyczne wspierające proces wydawniczy – zarówno czasopism, jak i publikacji zwartych. Działają również bazy czasopism, które pomagają rozpowszechniać już wydane publikacje.

Niezależnie od stosowanych narzędzi, należy pamiętać o zapewnieniu wysokiej jakości metadanych publikacji. Poprawnie i wyczerpująco opisane treści są dobrze widoczne w wyszukiwarkach internetowych. Jednocześnie ważne jest, by udostępniać poszczególne artykuły w oddzielnych plikach. Całe numery czasopisma dostępne w jednym pliku PDF, pozbawione szczegółowych danych bibliograficznych poszczególnych artykułów, w niewielkim stopniu pomagają publikacji stać się lepiej widoczną (pomimo że staje się ona dostępna w sposób otwarty).

Dobre metadane są jednym z warunków interoperacyjności, czyli zdolności do efektywnej współpracy między różnymi systemami. Dzięki niej treści udostępniane w systemach takich jak repozytoria czy platformy czasopism mogą być łatwo wyszukiwane w wielu innych miejscach w Internecie.

#### **4.4.1. Narzędzia informatyczne dostosowane do typu publikacji**

Poniżej przedstawiamy przykłady najpopularniejszych narzędzi: oprogramowania wspomagającego proces wydawniczy czasopisma naukowego, oprogramowania wspierającego wydawanie monografii oraz bazy czasopism ułatwiającej skuteczną dystrybucję periodyków.

##### **4.4.1.1. Open Journal Systems**

Open Journal Systems to system do zarządzania procesem wydawniczym czasopisma naukowego. Oprogramowanie jest udostępniane na wolnej licencji przez organizację Public Knowledge Project. Na świecie odnotowało już co najmniej 24 000 wdrożeń<sup>33</sup>. Ułatwia ono realizację wszystkich etapów procesu wydawniczego, od złożenia manuskryptu przez autora, przez recenzję zgodną z przyjętymi standardami (np. recenzja podwójnie ślepa), redakcję, aż po udostępnienie w Internecie. Oprogramowanie można dostosować do własnych potrzeb, zarówno jeśli chodzi o specyfikę procesu wydawniczego w danym czasopiśmie, jak i pod względem graficznym. OJS może służyć do wydawania jednego tytułu, jak również do tworzenia całych platform czasopism, na których znajduje się kilkadziesiąt lub więcej periodyków. Przykładem zastosowania OJS w Polsce jest Akademicka Platforma Czasopism Uniwersytetu Mikołaja Kopernika.

##### **4.4.1.2. Open Monograph Press**

Projektem bliźniaczym do OJS jest Open Monograph Press – bezpłatne oprogramowanie wspomagające proces wydawania monografii i ich upowszechniania w modelu otwartym. Również ono zostało przygotowane przez Public Knowledge Project i jest udostępniane na wolnej licencji. Przykładem polskiego wydawnictwa, które korzysta z tego rozwiązania, jest TMKarpiański Publisher. Dotychczas opublikowało m.in. trzy monografie naukowe. Lista przykładowych wdrożeń OMP na świecie jest dostępna na stronie Public Knowledge Project.

---

<sup>33</sup> <https://pkp.sfu.ca/ojs/ojs-usage/>

### 4.4.1.3. Bazy czasopism

Wydawca czasopisma może je udostępniać za pośrednictwem wyspecjalizowanych baz czasopism naukowych. Można wskazać trzy główne korzyści z takiego rozwiązania. Po pierwsze, dobrze prowadzone bazy zapewniają odpowiedni standard i wysoką jakość metadanych, dzięki czemu treści są dobrze widoczne w Internecie, w tym w popularnych wyszukiwarkach. Metadane wysokiej jakości umożliwiają automatyczną wymianę informacji między różnymi serwisami, tworzonymi zarówno na potrzeby upowszechniania publikacji, jak i sprawozdawczości. Po drugie, treści te będą widoczne dla użytkowników baz, którzy przeglądają zgromadzone w niej zasoby, a nie dotarliby do czasopisma, gdyby nie znajdowało się w bazie. Po trzecie, bazy zapewniają bezpieczne i długotrwałe przechowywanie treści.

Największą polską kolekcją otwartych czasopism jest [Biblioteka Nauki](#), w której skład wchodzi 5 baz czasopism, łącznie obejmujących wszystkie obszary wiedzy. W grudniu 2015 roku ponad 530 polskich czasopism naukowych udostępniało za pośrednictwem Biblioteki Nauki przynajmniej numery bieżące.

### 4.4.2. Metadane

W komunikacji naukowej wciąż wzrasta znaczenie wyszukiwarek, agregatorów i narzędzi do wymiany informacji. Nie do przecenienia jest więc rola metadanych, czyli usystematyzowanego opisu publikacji. Istnieją różne standardy metadanych. Popularny Dublin Core jest zalecany przez OpenAIRE<sup>34</sup>. Z kolei Google Scholar rekomenduje inne standardy, m.in. Highwire Press<sup>35</sup>. Uzasadnia to tym, że Dublin Core jest zbyt niejednoznaczny dla robotów indeksujących. Standardy metadanych nie wykluczają się – ten sam obiekt może zostać opisany równocześnie za pomocą metadanych w różnych formatach, o ile pozwala na to narzędzie (np. oprogramowanie bazy danych).

Rekomendujemy wydawcom przede wszystkim gruntowne zapoznanie się z problematyką metadanych i ich standardów. Decyzje o wyborze standardu metadanych należą do najważniejszych dla zapewnienia dobrej widoczności publikacji w Internecie. Metadane powinny być szczegółowe i jednoznaczne. Należy również zadbać o dostosowanie wybranego standardu do narzędzi służących udostępnianiu treści. W przypadku udostępniania różnych wersji publikacji zalecamy informowanie czytelnika, z którą wersją ma do czynienia. Jednocześnie metadane innych wersji niż ostateczna powinny zawierać link do wersji ostatecznej, optymalnie z pomocą DOI (nawet wtedy, kiedy wersja ostateczna jest zamknięta).

#### 4.4.2.1. DOI (Digital Object Identifier)

Ważnym elementem metadanych powinien być DOI (Digital Object Identifier), czyli cyfrowy identyfikator dokumentu elektronicznego. Dzięki posiadaniu DOI dana publikacja (np. artykuł lub monografia) będzie poprawnie identyfikowana niezależnie od ewentualnych zmian adresu URL czy jej metadanych, a także niezależnie od liczby miejsc, w których jest udostępniana.

Nadawanie publikacjom DOI jest korzystne niezależnie od tego, czy będą udostępniane w sposób otwarty, czy nie.

---

34 Por. [https://guidelines.openaire.eu/en/latest/literature/use\\_of\\_oai\\_dc.html](https://guidelines.openaire.eu/en/latest/literature/use_of_oai_dc.html).

35 <https://scholar.google.pl/intl/en/scholar/inclusion.html/>

Jest to usługa płatna świadczona przez agencje rejestrujące, których lista znajduje się na stronie DOI.

#### 4.4.3. Interoperacyjność

Interoperacyjność, czyli zdolność efektywnej współpracy różnych systemów, jest jednym z najważniejszych celów, jakie należy osiągnąć, budując narzędzia informatyczne służące otwartemu udostępnianiu treści naukowych. Zarys związanych z tym problemów z perspektywy repozytoriów został przedstawiony w dokumencie COAR, organizacji zrzeszającej otwarte repozytoria, ale większość poruszonych zagadnień odnosi się nie tylko do nich<sup>36</sup>.

Instytucja naukowa powinna uwzględnić potrzebę interoperacyjności, jeśli zdecyduje się tworzyć jakiegokolwiek własne narzędzia wspierające komunikację naukową lub korzystać z gotowych rozwiązań. Interoperacyjność ma znaczenie nie tylko w przypadku repozytoriów, ale również platform czasopism czy serwisów udostępniających monografie i innego typu publikacje. Zaniechania popełnione w tym względzie na początku mogą skutkować koniecznością dużego nakładu pracy i środków w celu ich naprawienia w przyszłości.

#### 4.4.4. Widoczność w Internecie

Naukowcy, szukając publikacji potrzebnych im w pracy badawczej, najczęściej zaczynają – a często również kończą – na popularnych wyszukiwarkach internetowych: Google i Google Scholar<sup>37</sup>. Wydawcy publikacji naukowych powinni ten fakt uwzględnić. Widoczność w Internecie jest niezbędnym uzupełnieniem otwartości. Publikacje, które są otwarte, ale nie są widoczne, w rzeczywistości nie wykorzystują potencjału nowych technologii.

Widoczność zależy od szeregu warunków<sup>38</sup>. Szczegółowe wskazówki dotyczące widoczności w odniesieniu do repozytoriów znajdują się w rozdziale piątym niniejszego raportu.

#### 4.4.5. Enhanced publications

Coraz większego znaczenia w środowisku cyfrowym nabierają tzw. enhanced publications, czyli publikacje wzbogacone o dodatkowe informacje. Przegląd różnych rodzajów takich publikacji znajduje się w przywoływanym już raporcie STM<sup>39</sup>. W szczególności wchodzi tu w grę połączenie artykułu z danymi badawczymi, na podstawie których ten artykuł powstał. Dane te mogą albo mieć postać surową, albo zostać w jakiś sposób zwizualizowane. Wizualizacje mogą z kolei zostać dołączone do pliku PDF lub być umieszczone wprost na stronie obok tekstu właściwego. Eksperymenty z tego rodzaju publikacjami przeprowadzają

---

36 <https://www.coar-repositories.org/activities/repository-interoperability/coar-interoperability-project/a-case-for-interoperability/>

37 Por. np. M. Kemman, M. Kleppe, S. Scagliola, 'Just Google It' [w:] C. Mills, M. Pidd, E. Ward, *Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities*, Sheffield: HRI Online Publications, 2014, <http://www.hrionline.ac.uk/openbook/chapter/dhc2012-kemman>. Choć naukowcy humaniści korzystają najczęściej z wyszukiwarki Google, to jednak większym zaufaniem darzą bazy takie jak JSTOR czy National Archief – por. rys. 5, s. 12 w cytowanym artykule.

38 Por.: <http://otwartanauka.pl/analysis/case-studies?id=945> oraz <http://otwartanauka.pl/analysis/rozwiązania/jak-poprawic-widoczność-rekordów-w-google-scholar>. Choć artykuł dotyczy przede wszystkim widoczności publikacji zdeponowanych w repozytoriach, część uwag ma zastosowanie również do innych sposobów obecności w Internecie.

39 *The STM Report*, 4th edition, s. 145–146.

zarówno wielcy wydawcy (choćby Elsevier), jak i nowe inicjatywy wydawnicze – np. periodyk eLife, rozwijający nowatorski format prezentacji artykułów naukowych, który nazwano Lens<sup>40</sup>.

#### 4.5. Czasopisma a monografie

Z punktu widzenia wydawcy publikacje naukowe dzielą się na dwie zasadnicze kategorie: monografie i periodyki. Warto zadać pytanie, czy postulaty otwartości odnoszą się w takim samym stopniu do nich obu. W dyskusjach o otwartym dostępie na świecie często można spotkać się z opinią, że łatwiej jest wdrożyć otwartość w odniesieniu do czasopism niż do monografii.

Z drugiej strony w sytuacji, gdy zarówno publikacja czasopisma, jak i monografii finansowana jest z zewnętrznej dotacji oraz nie wiąże się z oczekiwaniem znaczących przychodów ze sprzedaży, sytuacja periodyków i monografii staje się podobna.

Raport Stowarzyszenia STM wskazuje, że liczba monografii w otwartym dostępie jest wciąż stosunkowo mała. W humanistyce stosuje się model opóźnionego dostępu, który pozwala uzyskać pewien przychód ze sprzedaży publikacji. Model bazujący na opłatach za publikację jest rzadko spotykany z powodu wysokości tych opłat i braku dostępnych źródeł ich pokrycia. Szerzej przyjmuje się on w naukach ścisłych i przyrodniczych, gdzie w monografiach wieloautorskich jeden rozdział odpowiada artykułowi, przez co koszty rozkładają się na wiele osób lub instytucji<sup>41</sup>. Przykładowe oferty wydawców zostały zebrane w raporcie *Guide to Open Access Monograph Publishing for Arts, Humanities and Social Science Researchers*. Ceny zależą od długości tekstu. Mieszczą się w przedziale 3–11 tys. GBP<sup>42</sup>.

Katalog DOAB informuje o ponad 3200 recenzowanych książkach dostępnych w modelu otwartym, niezależnie od formy ich finansowania<sup>43</sup>.

Z kolei pisany z perspektywy Wielkiej Brytanii raport poświęcony wyłącznie monografiom w otwartym dostępie wskazuje, że wciąż nie wykształcił się sprawdzony system finansowania tego rodzaju publikacji. Dotychczasowe pilotaże nie przyniosły jeszcze wystarczającej ilości twardych danych, by móc dokonywać podsumowań<sup>44</sup>.

Elektroniczne formaty książek otwierają szerokie możliwości wzbogacania publikacji o funkcje, których książki papierowe nie mogły spełniać. Podobnie jak w przypadku artykułów określanych jako *enhanced publications* (patrz wyżej), książki elektroniczne mogą wykraczać poza pliki, takie jak PDF czy EPUB, oraz łączyć tradycyjną treść z danymi i ich wizualizacjami. Wokół cyfrowej książki można zbudować społeczność dyskutujących o niej naukowców. Przykładem udanego przedsięwzięcia tego typu jest *Debates in the Digital Humanities*.

---

40 <http://lens.elifesciences.org/about/>

41 *The STM Report*, 4th edition, s. 118.

42 E. Collins, C. Milloy, G. Stone, *Guide to Open Access Monograph Publishing for Arts, Humanities and Social Science Researchers*, OAPEN-UK, 2015, tab. 2, s. 18, <http://oapen-uk.jiscbooks.org/files/2015/07/Guide-to-open-access-monograph-publishing-for-researchers-final.pdf>.

43 <http://www.doabooks.org>

44 *Monographs and Open Access. A report to HEFCE*, Londyn 2015, s. 62 i n.

## 4.6. Publikacje bieżące a publikacje archiwalne

Instytucja naukowa zamierzająca wdrożyć otwarty dostęp powinna w pierwszej kolejności zwrócić uwagę na bieżące publikacje. W tym kontekście pożyteczne może być przeprowadzenie przeglądu obecnie stosowanych praktyk i obowiązujących regulacji. Może być też potrzebna modyfikacja części z nich, np. treści umów z autorami. Może się jednak również okazać, że w danej instytucji funkcjonują już praktyki i regulacje wystarczające, by rozpocząć wdrażanie otwartego dostępu.

Warto pamiętać również o publikacjach starszych. W takim przypadku zapewnienie otwartego dostępu jest pod pewnymi względami prostsze niż w sytuacji otwartego udostępniania nowych pozycji, jednak wiąże się też z pewnymi trudnościami. I tak jest ono prostsze, ponieważ komercyjny potencjał takich publikacji został już zrealizowany. Otwarty dostęp w przypadku starszych publikacji jest po prostu przywróceniem ich do obiegu z wykorzystaniem nowych kanałów dystrybucji. Zarazem jednak ich otwarte udostępnianie może wiązać się z trudnościami, z którymi nie trzeba mierzyć się w przypadku bieżących publikacji. Trudności te dzielą się na dwie kategorie: związane z prawami autorskimi oraz związane z koniecznością zmiany postaci publikacji (z wersji drukowanej na wersję elektroniczną).

Potrzeba otwartego udostępniania starszych publikacji różni się w zależności od dyscypliny nauki. W jednych (np. w naukach medycznych czy przyrodniczych) jest ona raczej niewielka, podczas gdy w innych (np. w naukach matematycznych lub humanistycznych) – bardzo duża.

### 4.6.1. Prawne aspekty udostępniania starszych publikacji

Czynności, jakie należy wykonać, aby legalnie udostępniać starsze publikacje, zależą od postanowień umów zawartych między wydawcą a autorem, a także od tego, kiedy zostały one zawarte. Nawet jeśli autor przeniósł wszystkie autorskie prawa majątkowe na wydawcę, ale uczynił to, zanim Internet zaczął funkcjonować jako pole eksploatacji (często przyjmuje się tu rok 1994 jako graniczny w Polsce; od tego czasu Internet w Polsce stał się dostępny nie tylko dla wybranych instytucji), to zgodnie z przepisami prawa autorskiego uprawnienie do udostępniania takiej publikacji w Internecie pozostaje przy autorze lub jego spadkobiercach. Jeżeli autor udzielił licencji tylko na publikację drukowaną, to również i w tym przypadku wydawca nie może sam decydować o udostępnieniu publikacji w Internecie. Oznacza to, że zapewnienie otwartego dostępu do starszych publikacji przez wydawnictwo wymaga wykonania znacznej pracy polegającej na uzyskaniu odpowiednich uprawnień<sup>45</sup>. Instytucja naukowa planująca udostępnić w modelu otwartym starsze publikacje może robić to albo wykonując pracę oddzielnie dla poszczególnych tytułów, albo przeprowadzając audyt stanu prawnego ogółu starszych publikacji. Taki audyt pozwoli zorientować się, jakie możliwości działania ma w tym zakresie dana instytucja.

### 4.6.2. Postać publikacji

W przypadku starszych publikacji instytucja zamierzająca udostępnić je w sposób otwarty może dysponować jedynie ich papierową wersją, a co za tym idzie – stanąć przed koniecznością ich digitalizacji. Na pewno ma to

---

45 K. Siewicz, *Otwarty dostęp do publikacji naukowych*, op. cit., s. 59–60.; S. Stanisławska-Kloc, *Polskie prawo autorskie dla bibliotek cyfrowych*, <http://www.bn.org.pl/download/document/1279799669.ppt>.

miejsce w przypadku publikacji powstałych w czasie, kiedy do składu nie wykorzystywano jeszcze komputerów. Taka potrzeba może zająć również wtedy, gdy w przypadku stosunkowo nowych publikacji z jakiegoś powodu nie zachowały się pliki z wersją do druku.

Standardy i dobre praktyki digitalizacji zostały wypracowane przez instytucje kultury – katalogi dobrych praktyk i zalecenia można znaleźć na stronie Narodowego Instytutu Muzealnictwa i Ochrony Zabytków<sup>46</sup>. Instytucje naukowe zamierzające udostępnić starsze publikacje powinny rozważyć sięgnięcie po pomoc podmiotów wyspecjalizowanych w przeprowadzaniu digitalizacji, która wymaga odpowiedniego sprzętu i kompetencji.

#### 4.7. Podsumowanie

Zachęcamy, by instytucje naukowe, które zamierzają wdrożyć otwarty dostęp do wydawanych przez siebie publikacji, jasno określiły priorytety tej działalności. Cele drugorzędne, możliwe do zrealizowania w krótkim czasie, nie powinny utrudniać realizacji celów zasadniczych, ale dających się osiągnąć dopiero w dalszej perspektywie. Wśród celów priorytetowych mogą znajdować się takie, jak wspieranie pracowników naukowych w budowaniu pozycji, prestiżu i spełnianiu wymagań instytucji finansujących badania, a także wspieranie instytucji naukowej we wzmacnianiu pozycji w środowisku akademickim i poza nim.

Niezależnie od pozostałych rozwiązań, instytucja powinna zapewnić autorom możliwość udostępniania prac w otwartych repozytoriach. Uregulowanie stosunków prawnoautorskich z autorami może przybrać różne formy, ale powinno mieć na celu odniesienie korzyści przez obie strony.

Skuteczne wykorzystanie nowych technologii w obszarze treści naukowych wymaga nie tylko ich udostępnienia w sposób otwarty, ale również zapewnienia ich jak najlepszej widoczności w Internecie. W związku z tym zachęcamy do starannego rozważenia technicznych aspektów udostępniania treści, w szczególności wyboru oprogramowania odpowiedniego do rodzaju publikacji i zapewniającego interoperacyjność, a także opracowania możliwie bogatych metadanych zgodnych ze światowymi standardami.

Zachęcamy instytucje do eksperymentowania z modelami otwartymi. Pewne rozwiązania można zastosować na zasadzie pilotażu, w odniesieniu do wybranych publikacji, a następnie ocenić ich efekty i podjąć decyzję co do ewentualnego wdrożenia. W szczególności władze instytucji powinny zapewnić warunki (materialne i formalne), by kierownictwo jednostek wydawniczych (wydawnictw) otrzymało większy zakres swobody pozwalający na wypróbowanie różnych rozwiązań.

---

46 <http://nimos.pl/pl/dzialalnosc/programy-wsparcia-finansowego/ochrona-i-cyfryzacja-dziedzictwa-kulturowego-1/katalogi-dobrych-praktyk-i-zalecenia-2015>



# Rozdział 5

## Repozytoria

W wydanym w 2015 roku przez Confederation of Open Access Repositories (COAR) dokumencie *COAR Roadmap – Future Directions for Repository Interoperability* wymieniono główne systemy tworzące obecnie obieg informacji naukowej. Znajdują się wśród nich repozytoria – zarówno instytucjonalne, jak i dziedzinowe<sup>47</sup>. COAR tradycyjnie wyróżnia trzy główne funkcje repozytoriów. Po pierwsze, stanowią one jedno ze źródeł publikacji naukowych, przez co kładą podwaliny pod system ich rozpowszechniania. Po drugie, odzwierciedlają rosnące zaangażowanie instytucji naukowych w zarządzanie dorobkiem swoich pracowników. Po trzecie w końcu, stanowią źródło informacji dla instytucji finansujących badania (w tym państwowych) o wielkości i wpływie rezultatów badań przez nie finansowanych<sup>48</sup>. Jeśli dodać do tego fakt, że od początku rozwoju ruchu otwartego dostępu repozytoria wskazywane są jako podstawowe narzędzie w jednym z głównych modeli jego wdrażania – zielonej drodze otwartego dostępu<sup>49</sup> – ich znaczenie okazuje się nie do przecenienia.

W raporcie *Otwarta nauka w Polsce 2014. Diagnoza* określiliśmy repozytorium jako „narzędzie informatyczne służące do deponowania, przechowywania i udostępniania w Internecie dorobku naukowego”. Repozytoria naukowe dzielimy na instytucjonalne, tj. udostępniające bieżący dorobek instytucji naukowej, oraz dziedzinowe, tj. udostępniające dorobek naukowy z określonej dziedziny (lub dziedzin) nauki<sup>50</sup>. W niniejszym rozdziale przedstawiamy rekomendacje dla tak rozumianych repozytoriów naukowych.

Większa część niniejszego rozdziału i najbardziej szczegółowe rekomendacje zostaną poświęcone kwestii widoczności repozytoriów w Internecie z punktu widzenia użytkownika. Jeśli poza funkcją przechowywania i archiwizacji dorobku naukowego repozytorium ma z równym powodzeniem pełnić funkcję udostępniania tego dorobku w sposób otwarty, to projektując je i prowadząc, należy mieć na uwadze zarówno kanały informacyjne, za pomocą których użytkownicy docierają do umieszczonych w repozytorium publikacji,

---

47 F. Summann i in., *COAR Roadmap – Future Directions for Repository Interoperability*.

48 K. Shearer, *Aligning Repository Networks and the Confederation of Open Access Repositories* (COAR), 2015. Plik PDF dostępny pod adresem: <http://ir.inflibnet.ac.in/bitstream/1944/1886/1/47.pdf>.

49 S. Harnad et al., *The access/impact problem and the green and gold roads to open access*, „Serials review” 2004, 30.4, s. 310–314. Plik PDF dostępny pod adresem: <http://www.ost.qc.ca/Portals/0/docs/articles/2004/science.pdf>.

50 *Otwarta Nauka w Polsce. Diagnoza*, red. J. Szprot, Wydawnictwa ICM, Warszawa 2014, s. 30.

jak i specyficzne wymagania, jakie trzeba spełnić, by publikacje były w owych kanałach obecne w liczbie reprezentatywnej dla wielkości zasobów danego repozytorium.

Statystyki wskazują, że większość użytkowników odnajduje materiały zgromadzone w repozytorium poprzez wyszukiwarki, a nie przez jego stronę główną<sup>51</sup>. Drugim z kolei kanałem dystrybucji są media społecznościowe. Jest to jedynie szczególny przypadek ogólnego trendu – z analogiczną sytuacją mamy do czynienia w przypadku serwisów z wiadomościami, w blogosferze itd. To cecha charakterystyczna dla dzisiejszego Internetu. Większość ruchu generują wyszukiwarki, zaś w drugiej kolejności media społecznościowe. To tam znajdują się potencjalni użytkownicy, którym należy wyjść naprzeciw. By to zrobić, trzeba najpierw znać mechanikę działania narzędzi, za pomocą których użytkownicy odnajdują zasoby w repozytoriach. Nie można też pominąć szerszego spojrzenia na repozytoria w kontekście systemu rozpowszechniania publikacji naukowych w Internecie – dopiero na tym tle będziemy widzieć, jakie zadania stoją przed repozytoriami.

W pierwszej części niniejszego rozdziału postaramy się więc zarysować perspektywę internetowego systemu udostępniania publikacji naukowych, którego repozytoria są częścią. Druga część poświęcona będzie europejskiej infrastrukturze otwartego dostępu, tworzonej w ramach projektu OpenAIRE. W części trzeciej wykażemy, że wyszukiwarki internetowe (akademickie i ogólnego przeznaczenia) są głównym „punktem wejścia” dla użytkowników repozytoriów, po czym opiszemy podstawy działania takich wyszukiwarek oraz główne problemy stojące przed repozytoriami naukowymi stawiającymi sobie za cel dobrą widoczność w wyszukiwarkach. Omówimy też zasady działania podstawowych narzędzi, za pomocą których wyszukiwarki gromadzą dane o indeksowanych stronach. W czwartej części wskażemy podstawowe trudności związane z ekstrakcją metadanych z tekstu, by podkreślić kluczowe znaczenie dostarczania metadanych gotowych do maszynowego odczytania. Następnie przedstawimy rekomendacje, które będą oparte przede wszystkim o dotychczasowy dorobek COAR, wytyczne OpenAIRE oraz porady dotyczące SEO dla repozytoriów pojawiające się w różnych źródłach<sup>52</sup>. Rekomendacje będą dostosowane do lokalnych uwarunkowań o tyle, o ile sytuacja polskich repozytoriów w znaczący sposób różni się pod danym względem od przeciętnej sytuacji światowej.

## 5.1. Repozytoria jako część systemu udostępniania publikacji naukowych w Internecie

Repozytorium publikacji naukowych to jedno z narzędzi, dzięki którym dana publikacja „jest dostępna w Internecie”. Sama dostępność to jednak oczywiście nie wszystko. Kluczowe znaczenie dla szukającego publikacji użytkownika repozytorium ma w tym kontekście widoczność, czyli informacja o tym, że dana publikacja – w szczególności dany artykuł – jest dostępna. W Internecie artykuły naukowe konkurują ze sobą pod względem widoczności na dwa niezależne od siebie sposoby: jako artykuły naukowe i jako materiały dostępne w sieci. W obu przypadkach użytkownik końcowy postawiony jest w sytuacji natłoku

51 Z raportów ogólnodostępnych za pośrednictwem Internetu zob. np. B. Matthies, *Digital Commons Annual Report 2011*. Butler University Library, 2011.

52 Por.: K. Arlitsch, P. O'Brien, *Improving the visibility and use of digital repositories through SEO*. American Library Association, 2013; D. Dapra (Google Scholar), *Indexing Repositories: Pitfalls and Best Strategies*, 2015; bepress, *Search Engine Optimization: How to Attract More Visitors to Your Repository, Digital Commons Reference Material and User Guides*. Paper 35, 2012.

informacji, które musi wstępnie przefiltrować, aby następnie móc z nich skorzystać. W pierwszym przypadku rolę filtrów spełniają m.in. różnorakie kryteria i narzędzia ewaluacji dorobku naukowego. W drugim użytkownik do pewnego stopnia sam spełnia funkcję takiego filtra – to on wybiera odwiedzane przez siebie strony internetowe, buduje sieć kontaktów w mediach społecznościowych itp. Poza tym wycinkiem sieci korzysta zaś z wyszukiwarek internetowych. Jak się przekonamy, podstawowe zasady działania narzędzi ewaluacyjnych dorobku naukowego oraz podstawowe algorytmy wykorzystywane przez wyszukiwarki internetowe mają pewien wspólny rdzeń. W obu aspektach „walki o użytkownika” repozytoria dysponują – lub mogą i powinny dysponować – szeregiem narzędzi pozwalających na przekazanie mu informacji na temat istnienia i zawartości danego zasobu.

Jak wspomnieliśmy, repozytoria dzieli się zazwyczaj na instytucjonalne i dziedzinowe. Co je odróżnia? Pod względem technicznym praktycznie nic – oprogramowanie repozytoryjne w obu przypadkach jest takie samo. Różnica dotyczy charakteru zasobów w obu typach repozytoriów, ale ma to znaczenie tylko w takim stopniu, w jakim użytkownik korzysta z nich bezpośrednio na stronie internetowej danego repozytorium. Jeżeli potraktujemy repozytoria jako źródło zasobów dla korzystających z nich systemów agregujących oraz wyszukiwarek, różnica ta okazuje się mniej istotna. Istotną różnicą pozostaje natomiast model działania i utrzymania obu typów repozytoriów. W przypadku repozytoriów instytucjonalnych sprawa jest stosunkowo prosta: za ich sprawne działanie oraz bezpieczeństwo i trwałość zgromadzonych w nich zasobów odpowiedzialne są instytucje, których dorobek jest w nich udostępniany. W przypadku repozytoriów dziedzinowych brak podobnie oczywistego rozwiązania. W szczególności zapewnienie ciągłości działania oraz rozwoju mniejszych i słabiej rozpoznawalnych repozytoriów dziedzinowych może okazać się poważnym wyzwaniem.

W przywoływanym wcześniej dokumencie *COAR Roadmap* repozytoria (w ogromnej większości instytucjonalne<sup>53</sup>) stanowią podstawową warstwę systemu rozpowszechniania publikacji i danych badawczych przez Internet. Ich zadaniem jest archiwizacja i udostępnianie publikacji wraz z dobrze opracowanymi metadanymi. Są one oczywiście dostępne dla użytkownika końcowego, jednak w modelowym ujęciu większość użytkowników końcowych mogłaby stykać się z publikacjami zdeponowanymi w repozytoriach z pomocą kilku warstw (lub serwisów) pośrednich. Dane z repozytoriów mogą być dostępne dla zewnętrznych serwisów, które filtrują dla użytkowników końcowych dostępne w nich publikacje w taki sposób, by po wyborze odpowiedniego narzędzia filtrującego – względnie po odpowiednim ustawieniu opcji w tymże – każdy użytkownik mógł liczyć na odfiltrowanie interesujących go publikacji. Idea działania systemu naszkicowanego przez COAR jest bardzo podobna, z tym że zamiast filtrować zasoby jednego repozytorium, odpowiednie serwisy miałyby do czynienia z zebranymi razem zasobami wszystkich repozytoriów zintegrowanych z systemem. Oczywiście takim agregowaniem publikacji i danych zajmowałyby się inne serwisy – agregatory. Oddzielnym zadaniem byłoby usunięcie duplikatów z zasobów (tak by użytkownik końcowy w wyniku zagregowania nie otrzymywał wszystkich wersji jednego artykułu widocznych jako oddzielne publikacje), zarządzanie metadanymi itp.

53 Por. statystyki OpenDOAR: 84,2% wszystkich zarejestrowanych repozytoriów to repozytoria instytucjonalne. Dane z dnia 03.09.2015 r. (wykres z aktualnymi danymi dostępny pod adresem: <http://tinyurl.com/typyRepoWykresOpenDOAR>).

Obecnie mamy już do czynienia z serwisami, które spełniają część zadań niezbędnych do realizacji tej wizji. PubMed Central w Ameryce Północnej, La Referencia w Ameryce Łacińskiej czy też OpenAIRE w Europie w jednym miejscu dostarczają użytkownikom informacji zebranych z różnych źródeł. OpenAIRE dodatkowo opracowuje i uzupełnia metadane, a także odgrywa kluczową rolę w procesie ich ujednolicania poprzez system walidacji i rejestracji repozytoriów oraz regularnie aktualizowane rekomendacje dotyczące użycia OAI-PMH oraz Dublin Core.

Do pewnego stopnia rolę agregatorów pełnią również internetowe wyszukiwarki akademickie. Google Scholar deduplikuje zagregowane publikacje, częściowo wymusza też stosowanie jednolitych metadanych na instytucjach, którym zależy na widoczności w tym serwisie. Świadczy też dodatkowe usługi statystyczne użytkownikom końcowym, np. wylicza cytowania i indeks Hirscha zarówno czasopism<sup>54</sup>, jak i poszczególnym zarejestrowanym autorom. Z drugiej strony serwis ten nie ingeruje w żaden sposób w zagregowane metadane, nie oddziela też publikacji naukowych od pozostałych w sposób inny niż zautomatyzowany.

Może się wydawać, że w roli mało widocznych dla użytkownika końcowego dostawców publikacji i metadanych do innych serwisów repozytoria mogą pozostać nieznane, zaś ich rola – niedoceniana. Trzeba jednak pamiętać, że – po pierwsze – zręby systemu rozpowszechniania treści naukowych już teraz działają w ten sposób. Google Scholar, dla przykładu, przekierowuje użytkowników bezpośrednio do plików PDF w sposób, który może nawet nie zostać uwzględniony w podstawowych statystykach wejść do repozytorium. Po drugie zaś, instytucje prowadzące repozytoria tak czy inaczej powinny prowadzić kompleksowe działania informacyjne i promocyjne (więcej o takich działaniach piszemy w rozdziale szóstym).

## 5.2. Europejska infrastruktura otwartego dostępu OpenAIRE

Seria projektów OpenAIRE rozpoczęła się w roku 2009 jako kontynuacja wcześniejszych projektów DRIVER i DRIVER II. Złożyły się na nią OpenAIRE (2009–2011, FP7), OpenAIREplus (2011–2014, FP7) i OpenAIRE2020 (2015–2018, H2020). Początkowo celem projektu było przede wszystkim wsparcie realizacji Pilotażu Otwartego Dostępu Komisji Europejskiej w ramach Siódmego Programu Ramowego (7PR), ale stopniowo jego zakres rozszerzał się w kierunku budowy kompleksowej europejskiej infrastruktury otwartego dostępu i otwartych danych. Dziś poza agregacją dostępnego w sposób otwarty europejskiego dorobku naukowego OpenAIRE pomaga m.in. w monitorowaniu wyników projektów realizowanych w ramach programu Horyzont 2020, wspiera realizację prowadzonego przez Komisję Europejską Pilotażu Otwartych Danych, współpracuje z instytucjami finansującymi badania, bada nowe formy komunikacji naukowej i wspiera wdrażanie otwartych modeli przy pomocy sieci Krajowych Biur Otwartego Dostępu (NOADs). Z punktu widzenia e-infrastruktury, OpenAIRE działa na terenie UE jako główna siła centralizująca w „modelu rozproszonym” obejmującym m.in. repozytoria. Promuje również ujednolicenie metadanych (zwłaszcza za pomocą protokołu OAI-PMH i formatu Dublin Core), prowadząc rejestrację repozytoriów. Każde repozytorium, które chce dołączyć do sieci

---

54 Por. np. [https://scholar.google.com/citations?view\\_op=top\\_venues](https://scholar.google.com/citations?view_op=top_venues).

OpenAIRE, musi być zgodne ze standardami *OpenAIRE Guidelines*. Zgodność tę można sprawdzić z pomocą walidatora on-line.

OpenAIRE odgrywa więc kluczową rolę w repozytoryjnej infrastrukturze Europy, a w perspektywie także na świecie – chociażby poprzez współpracę z La Referencia. Z tego względu rekomendujemy instytucjom prowadzącym repozytoria ścisłą współpracę z OpenAIRE.

## 5.3. Wyszukiwarki a treści naukowe w Internecie

### 5.3.1. Ogólna charakterystyka wyszukiwarek internetowych

Wyszukiwarka internetowa jest narzędziem pozwalającym użytkownikowi na odnajdywanie interesujących go stron poprzez wpisanie słów kluczowych wyrażających zakres jego zainteresowań. Na ich podstawie wyszukiwarka przeszukuje posiadany indeks stron znajdujących się w Internecie i za pomocą odpowiednich algorytmów stara się wybrać te, które z największym prawdopodobieństwem będą zawierać interesujące go treści. Bardzo ważnym czynnikiem jest w tym procesie czas. Wyszukiwarka jest tylko środkiem, który powinien zaprowadzić użytkownika do celu możliwie łatwo i szybko. Google, aktualnie niekwestionowany lider na rynku wyszukiwarek, na tyle ceni czas reakcji swojej wyszukiwarki, by prezentować go użytkownikom.

Krótki spodziewany czas reakcji oznacza oczywiście, że wyszukiwarka nie przeszukuje Internetu w czasie zapytania użytkownika. Korzysta z gotowego już wcześniej indeksu. Wszystkie strony zapisane w indeksie stara się ona ocenić w zależności od spodziewanej atrakcyjności ich treści dla czytelnika. Na serwerach danego serwisu w każdym momencie działa wiele instancji różnorodnych robotów indeksujących. Możliwości obliczeniowych serwerów nie są oczywiście nieograniczone. Jeśli crawler zapętlony się w strukturze danej strony internetowej, musi zbyt długo czekać na odpowiedź jej serwera lub jego algorytmy heurystyczne wskazują, że strona jest mało ciekawa dla użytkownika, wówczas dalsze przeznaczanie zasobu obliczeniowego na indeksowanie danej strony przestaje być w interesie serwisu<sup>55</sup>.

Główna idea oceny strony internetowej pod kątem atrakcyjności dla użytkowników podobna jest w swych najbardziej fundamentalnych zasadach do idei Impact Factor – wymiernym skutkiem zainteresowania użytkowników są linki prowadzące do strony. Im więcej linków prowadzi do strony, tym wyższa punktacja; rośnie ona również wtedy, gdy strony, z których wychodzą linki, same plasują się wysoko w tej samej konkurencji. Są to główne założenia PageRank, czyli serca systemu oceniającego stosowanego przez Google. Ocena końcowa zależy również od dużej liczby czynników (tylko część z nich jest znana – całość „przepisu” stanowi tajemnicę handlową firmy Google), które w teorii powiązane są z jakością użytkowania strony. Im wyżej w hierarchii oceny algorytmów heurystycznych Google znajduje się dana strona, tym wyżej wyświetli się ona na stronie prezentującej wyniki wyszukiwania.

Ponieważ główne założenia algorytmu oceniającego są znane, w naturalny sposób każdy, komu zależy na widoczności własnej strony internetowej, stara się jak najlepiej wypaść w tym rankingu. Tak powstała

---

55 Por. M. Starczewski, T. Lewandowski, *Po czym rozpoznać dobre repozytorium?*, „Biblioteka i Edukacja” 2014, nr 6, s. 8.

dziedzina wiedzy i praktyki znana jako SEO (*Search Engine Optimization*, czyli optymalizacja ze względu na wyszukiwarki internetowe). Gra toczy się o wysoką stawkę – badania pokazują, że w praktyce jedynie serwisy pojawiające się na pierwszej stronie wyników wyszukiwania mają szansę na pożądaną poziom ruchu internetowego<sup>56</sup>.

Na potrzeby niniejszego raportu skoncentrujemy się głównie na programach, z pomocą których wyszukiwarki internetowe konstruują i utrzymują swe indeksy. Programy te czytają źródło strony internetowej (HTML), identyfikują kluczowe z punktu widzenia wyszukiwarki informacje oraz – co stanowi podstawę wyliczenia PageRank – linki prowadzące do innych domen i stron w Internecie. Programy indeksujące wyszukiwarek akademickich dokonują poza tym dodatkowego parsowania<sup>57</sup> i – jeśli jest to potrzebne – ekstrakcji<sup>58</sup> w celu pozyskania odpowiednich danych bibliograficznych. Następnie programy indeksujące podążają za zidentyfikowanymi linkami i odczytują strony, do których prowadziły linki. Od takiego sposobu działania – podążania za poszczególnymi nitkami ogromnej sieci – programy takie zostały nazwane „crawlerami” (z ang. – pełzacze) czy też „pajęczkami”. W niniejszym tekście używamy również określeń „program indeksujący”, „robot indeksujący” oraz innych zamienników.

Informacje na temat crawlerów otrzymaliśmy od Google Scholar, jednak są one reprezentatywne dla robotów indeksujących wszystkich wyszukiwarek akademickich. Wiadomo przy tym, że obecnie Google Scholar zajmuje w tym obszarze pozycję dominującą. Wydaje się, że standardy rekomendowane przez Google Scholar są na tyle zdroworozsądkowe, że miałyby dużą szansę przyjęcia się nawet wtedy, gdyby GS nie rekomendowało ich stosowania.

### 5.3.2. Specyfika repozytoriów jako obiektów indeksowanych przez wyszukiwarki internetowe

Misją repozytoriów jest m.in. rozpowszechnianie dorobku naukowego, dlatego powinny one starać się przyciągnąć jak największą liczbę czytelników. Ponieważ zaś medium ich działalności stanowi Internet, dążenie to powinno być realizowane przez umiejętną optymalizację pod kątem wyszukiwarek internetowych.

We wspomnianym już raporcie biblioteki Butler University za 2011 rok<sup>59</sup> znajduje się informacja, że ponad połowa ruchu internetowego (53%) tej biblioteki generowana była przez wyszukiwarki, z czego ogromna większość (49% całości) przypadała na narzędzia firmy Google (wyszukiwarki ogólnego przeznaczenia oraz Google Scholar). Można zadać pytanie, jak rozkładają się wejścia na strony repozytoriów między tymi dwoma narzędziami Google. Dla przykładu, w przypadku [Repozytorium CeON](#) w pierwszej połowie 2015 roku na ok. 156 000 wejść z serwisu Google, ok. 22 500 pochodziło z Google Scholar.

---

56 Por. np. badania agencji Chitika z 2013 roku (Chitika Insights, *The Value of Google Result Positioning*, 2013, s. 6– 7): pierwsza strona wyników Google generuje wg nich 91,5% całego ruchu generowanego przez tę wyszukiwarkę. 33% ruchu z pierwszej strony jest z kolei generowane wyłącznie przez pierwszy wynik wyszukiwania. Podobne rezultaty wykazały badania z 2010 roku tej samej agencji. Skróconą wersję raportu można przeczytać pod adresem: <https://chitika.com/google-positioning-value>.

57 W tekście niniejszego rozdziału odróżniamy parsowanie od ekstrakcji. Oba terminy oznaczają proces uzyskiwania przez program informacji z tekstu, jednak o ile „parsowanie” oznacza pozyskiwanie informacji z tekstu uprzednio w tym celu przygotowanego, o tyle „ekstrakcja” będzie oznaczała proces odczytywania danych z tekstu w zamyśle skierowanego raczej do ludzkiego odbiorcy.

58 Patrz sekcje 5.4.1., 5.4.2. oraz 5.4.3.

59 B. Mathies, *Digital Commons Annual Report 2011*. Butler University Library, 2011.

Skoro wyszukiwarki ogólnego przeznaczenia generują większy ruch niż wyszukiwarki akademickie, to czy dla uzyskania lepszej widoczności zasobów zdeponowanych w repozytorium nie wystarczy zatroszczyć się jedynie (lub głównie) o SEO względem wyszukiwarek ogólnego przeznaczenia?

Odpowiedź na to pytanie jest negatywna z co najmniej dwóch powodów. Po pierwsze, specyfika działania wyszukiwarek akademickich sprawia, że najprawdopodobniej większość generowanego przez nie ruchu nie będzie się pojawiała w podstawowych statystykach repozytorium<sup>60</sup>. Po drugie, obecność publikacji w indeksach wyszukiwarek akademickich jest bardzo ważna z punktu widzenia naukowców posiadających konto w serwisie Google Scholar, który automatycznie prowadzi rejestr publikacji danego naukowca i ich cytowań oraz na tej podstawie oblicza indeks Hirscha. Ten współczynnik odgrywa coraz większą rolę w zarządzaniu przez pracowników własnym wizerunkiem naukowym oraz w ich karierze. Każde repozytorium powinno rozpoznać tę potrzebę naukowców i dołożyć starań, by została ona spełniona.

Kenning Arlitsch i Patrick O'Brien w książce *Improving Visibility and Use of Digital Repositories through SEO*, poświęconej SEO dla repozytoriów, stawiają tezę, że wśród obiektów indeksowanych przez wyszukiwarki internetowe repozytoria stanowią bardzo charakterystyczną grupę. Następnie starają się odpowiedzieć na pytanie, co przede wszystkim wyróżnia repozytoria. Według nich „sercem” repozytorium jest baza danych. Podstrony serwisu repozytoryjnego, na których użytkownik przegląda poszczególne rekordy, są konstruowane dynamicznie<sup>61</sup>. Użytkownik za pośrednictwem przeglądarki internetowej wysyła żądanie do serwera, ten zaś na podstawie zawartych w repozytorium metadanych konstruuje odpowiednią stronę w oparciu o szablon zawarty w systemie zarządzania treścią (CMS) danego serwisu repozytoryjnego. Podstrona z danymi bibliograficznymi oraz linkiem do odpowiedniego obiektu cyfrowego nie istnieje zatem, zanim użytkownik nie wykona zapytania<sup>62</sup>.

Wydajność strony repozytorium jest bardzo ważna z punktu widzenia widoczności zasobów w Internecie. Jeśli kolejne strony z rekordami będą konstruowane zbyt wolno, crawler przerwie indeksowanie danego repozytorium. Celem wyszukiwarki jest przecież zaindeksowanie – przy skończonych zasobach – maksymalnej liczby optymalnych jakościowo rekordów. Jeśli więc algorytm uzna, że przy pomocy zasobów zużywanych właśnie przez crawler będzie mógł w innym miejscu zaindeksować większą liczbę stron, to po prostu przeniesie tam indeksowanie. Kluczową składową tej wydajności może okazać się odpowiednia architektura repozytoryjnej bazy danych.

---

60 Chodzi o to, że wyszukiwarki akademickie zazwyczaj odsyłają czytelników bezpośrednio do pliku PDF. W związku z tym pobrania wykonane za pośrednictwem wyszukiwarek akademickich nie będą się pojawiały w statystykach dotyczących pobrań ze stron repozytoriów – o ile nie wprowadzi się odpowiednich modyfikacji. Przykład takiej modyfikacji dla Google Analytics Tools podają K. Arlitsch i P. O'Brien w *Improving the visibility and use of digital repositories through SEO*, American Library Association, 2013, s. 49.

61 Warto zaznaczyć, że dynamiczne tworzenie stron i podstron przez CMS na podstawie bazy danych nie jest charakterystyczne wyłącznie dla repozytoriów. Od dłuższego czasu jest to model dominujący wśród stron internetowych. Serwisy wykorzystujące np. oprogramowanie WordPress czy Joomla! działają w oparciu o identyczne zasady. W każdym przypadku czas ładowania strony będzie zależał w dużej mierze od czasu reakcji bazy danych.

62 K. Arlitsch, P. O'Brien, *Improving...*, op. cit., s. 29–30.

## 5.4. Metadane

Jak wspomnieliśmy powyżej, za konstrukcję indeksu stron internetowych używanego przez wyszukiwarkę internetową odpowiadają crawlery. Crawlery wyszukiwarek akademickich oprócz danych na temat samej strony internetowej zbierają również dane dotyczące publikacji naukowych. Dane o publikacji to szczególnie podzbiór metadanych, tj. danych (czy informacji) na temat danych. Niniejszy podrozdział poświęcimy omówieniu tego, w jaki sposób crawler wyszukiwarki zbiera metadane oraz jak powinny być one skonstruowane i gdzie umieszczone, by ich kolekcjonowanie było dla robotów indeksujących jak najłatwiejsze. Im łatwiej crawlerowi odczytać metadane, tym – statystycznie rzecz biorąc – więcej pozycji z repozytorium zostanie uwzględnione w indeksach wyszukiwarki.

### 5.4.1. HTML

Przeglądając strony internetowe, crawler „widzi” je inaczej niż zwykły użytkownik (czy raczej jego przeglądarka internetowa). Najważniejsze różnice to<sup>63</sup>:

1. brak możliwości przeglądania informacji zawartych w plikach graficznych, audio czy wideo<sup>64</sup>;
2. brak możliwości wypełniania wszelkich dostępnych na stronie formularzy<sup>65</sup>;
3. brak możliwości obsługi metody przesyłania danych POST<sup>66</sup>;
4. brak możliwości obsługi JavaScript<sup>67</sup>.

Różnice te można podsumować stwierdzeniem, że roboty indeksujące analizują kod HTML przeglądanej strony. Kod CSS, odpowiedzialny za rozmieszczenie zakodowanych w HTML elementów na ekranie, jest tu oczywiście nieistotny.

Język HTML (czyli *HyperText Markup Language*) jest językiem znaczników. W tym miejscu warto przypomnieć sobie typowy wygląd strony internetowej sprzed 2000 roku – albo jeszcze lepiej – z początków Internetu (rys. 1).

---

63 Por. *Jak poprawić widoczność rekordów w Google Scholar*, <https://otwartanauka.pl/analysis/rozwiązania/jak-poprawic-widocznośc-rekordów-w-google-scholar>.

64 Odczytują jednak oczywiście metadane powiązane z daną grafiką, filmem czy nagraniem dźwiękowym. Z punktu widzenia repozytoriów różnica ta jest ważna np. ze względu na to, że nawet wyspecjalizowana wyszukiwarka akademicka z zasobami dedykowanymi specjalnie plikom PDF nie odczyta tekstu, jeśli do pliku nie będzie dołączona warstwa tekstowa, gdyż bez niej PDF jest tylko grafiką. Podobnie w niektórych systemach zarządzania treścią stosowanych przez wydawców czasopism naukowych użytkownikowi prezentowana jest publikacja w formie serii plików graficznych. Człowiek odczyta treść udostępnioną w ten sposób, ale maszyna – nie. Obrazy czy dźwięki można oczywiście maszynowo zamienić na tekst z pomocą programów typu OCR (*Optical Character Recognition*) oraz *speech-to-text*, jednak w większości przypadków sama wyszukiwarka nie wykona tych działań przygotowawczych; zakłada się, że zostały wykonane przez obsługę repozytorium.

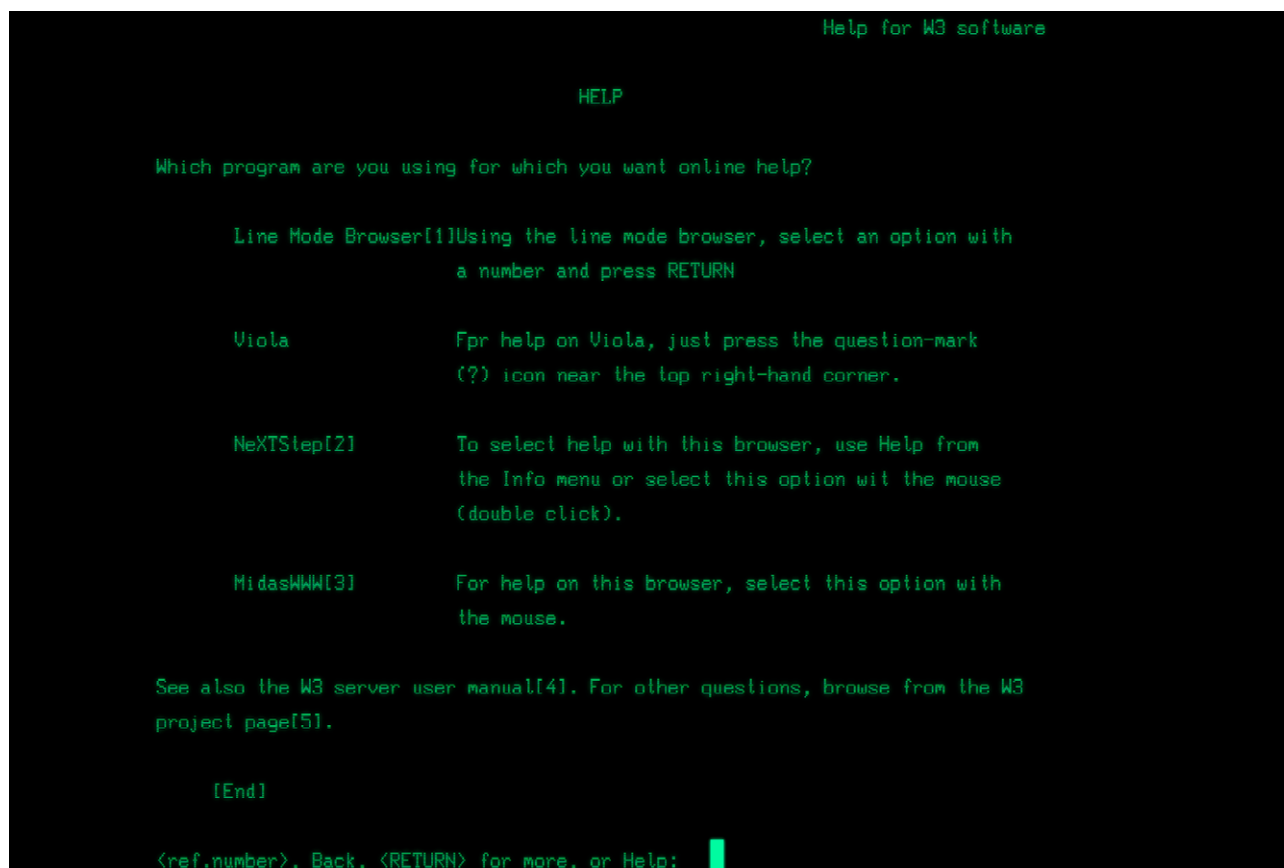
65 D. Dapra, *Indexing Repositories...*, s. 15–16, *Jak poprawić widoczność...*, op. cit.

66 D. Dapra, op. cit., *Jak poprawić widoczność...*, op. cit.

67 D. Dapra, op. cit., *Jak poprawić widoczność...*, op. cit.



Rysunek 1. Pierwsza działająca strona WWW info.cern.ch odtworzona dzięki dostępnemu on-line symulatorowi przeglądarki line-mode



Patrząc na nią, szybko uświadomimy sobie, że strona internetowa to przede wszystkim tekst. Zasadniczo HTML składa się z tekstu, którego fragmenty zostały otagowane w określony sposób. Przykładem tagu jest następujący ciąg znaków: `<a href="http://foo.bar">kliknij w link</a>`. „Kliknij w link” jest otagowanym tekstem. Elementy `<a>` oraz `</a>` wyznaczają początek i koniec tagu (niektóre tagi to tzw. tagi puste<sup>68</sup>, które nie obejmują tekstu, w związku z czym nie posiadają zamknięcia typu `</x>` – do takich tagów należą np. `<img>` odpowiedzialny za wklejanie w stronę plików z obrazami czy `<meta>`, który szerzej omówimy za chwilę). Z kolei `href` to atrybut tagu – w tym przypadku jest to skrót od „hyperlink reference” i oznacza stronę, do której link ma prowadzić. `"http://foo.bar/"` to wartość atrybutu `href`, = oznacza przypisanie wartości do atrybutu. Bywają również prostsze tagi, niezawierające atrybutów, np. `<b>tekst</b>` oznacza pogrubienie (wyboldowanie) tekstu. Wszystkie tagi muszą być zawarte w głównym tagu: `<html>(tu źródło strony)</html>`, aby nie zostały zignorowane przez przeglądarkę<sup>69</sup>.

Strona HTML (czyli to, co jest widoczne jako tzw. źródło strony<sup>70</sup>) składa się z dwóch części: `<head>` („głowy”) i `<body>` („ciała”). W sekcji `<body>` zasadniczo zakodowane zostają wszystkie elementy widoczne dla użytkownika na ekranie. W sekcji `<head>` znajdują się wszystkie inne elementy odpowiedzialne

68 Tzw. *void element*, <http://www.w3.org/TR/html-markup/syntax.html#void-element>.

69 Wszystkie informacje o podstawach języka HTML można znaleźć np. na stronie w3schools: <http://www.w3schools.com/html/>

70 We wszystkich popularnych przeglądarkach w menu kontekstowym dostępnym pod prawym klawiszem myszy znajduje się opcja „wyświetl źródło strony” lub równoważna. Źródło HTML można też przeglądać (i edytować) w każdym programie typu Developer Tools (Chrome Developer Tools dla Google Chrome, Firebug dla Firefox itp.).

za ich układ oraz zawierające dane o stronie dostępne np. dla wyszukiwarek internetowych (takie jak słowa kluczowe, opis zawartości strony, jej tytuł itp.). Skrypty Javascript też najczęściej umieszczane są w sekcji <head><sup>71</sup>. Tam też jest domyślnie miejsce tagów <meta><sup>72</sup>, w których zawarte są metadane dotyczące dokumentu HTML, przeznaczone jednak nie dla użytkownika przeglądarki, ale do parsowania maszynowego. Jako takie są najwłaściwszym miejscem na umieszczenie metadanych dotyczących również zasobu repozytoryjnego – na stronie zawierającej rekord bibliograficzny zasobu oraz ewentualnie link do zawierającego go pliku. Tag <meta> należy do opisanych wyżej tagów pustych, tj. nie obejmuje żadnego tekstu (przez co nie ma znacznika końcowego </meta>), jego zawartość podana jest w formie atrybutów z przypisanymi wartościami. I tak np. tag <meta charset="UTF-8"> określa kodowanie znaków dla dokumentu HTML. Wszystkie standardy metadanych dla publikacji i danych naukowych wykorzystują ten właśnie tag. Ma on wtedy zazwyczaj dwa atrybuty: "name" i "content". Wartość pierwszego atrybutu wyznacza rodzaj metadanych, wartość drugiego – ich treść. Poniżej widzimy kilka przykładów tagów <meta> zawierających metadane dotyczące publikacji:

```
<meta name="citation_author" content="Jan Kowalski">
<meta name="dc.creator" content="Maria Antonina">
<meta name="citation_date" content="1979">
<meta name="citation_lastpage" content="42">
```

Poszczególne standardy metadanych regulują możliwe wartości atrybutów "name". Roboty indeksujące rozpoznają standardy i znają zestaw reguł wyznaczający każdy z nich, dzięki czemu informacje zawarte w ten sposób w pliku HTML są bardzo łatwe do automatycznego odczytania. Jeśli repozytorium umieści w pliku dobrej jakości metadane, nie będą one wymagały niemal żadnej dalszej obróbki ani poprawiania. W przypadku agregatorów takich jak OpenAIRE, oszczędzi to pracy nad metadanymi i przyspieszy widoczność zasobów zdeponowanych w repozytorium w serwisach związanych z agregatorem. W przypadku partnerów takich jak Google Scholar, dobrej jakości metadane mogą zdecydować o tym, czy użytkownik końcowy znajdzie dany zasób, zaś w skrajnych przypadkach – w ogóle o umieszczeniu zasobu w indeksie.

#### 5.4.2. Dane bibliograficzne w rekordzie repozytorium

Pamiętając o podziale pliku HTML na sekcje head i body, jak również dysponując podstawowymi informacjami na temat tagów HTML (szczególnie tagów <meta>), możemy precyzyjnie wskazać różnicę pomiędzy informacjami bibliograficznymi zawartymi w pliku HTML dla robotów indeksujących a informacjami widocznymi dla użytkownika podczas przeglądania rekordów repozytorium. Ta sama informacja (np. o nazwisku autora) jest w tym samym dokumencie HTML podana w dwóch niezależnych miejscach.

---

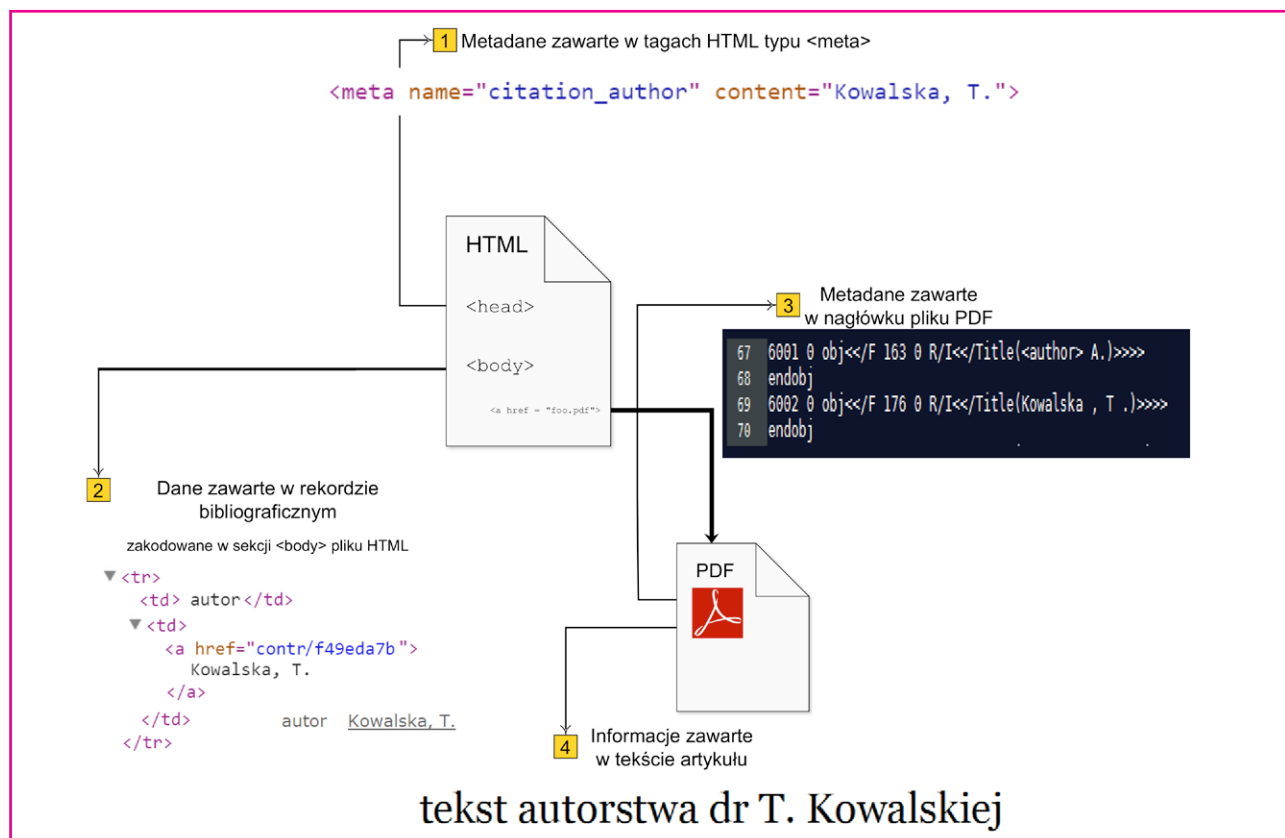
71 Zdarzają się wyjątki od tej zasady. Większość przeglądarek działa w ten sposób, że podczas „ładowania” sekcji <head> użytkownik nie może wykonać na stronie żadnej akcji, widzialne elementy zaś pojawiają się w miarę ładowania poszczególnych elementów z sekcji <body>. Jeśli więc mamy na stronie wyjątkowo „ciężki” Javascript, którego działanie nie jest kluczowe dla poprawnego załadowania któregoś z widzialnych/interaktywnych elementów strony, dobrą praktyką jest umieszczanie tego skryptu na końcu sekcji <body>. To złamanie ogólnej zasady użytkownika sekcji <body> i <head> ma bardzo ważny dla użytkowników skutek: użytkownik może wejść w interakcję ze stroną jeszcze zanim „ciężki” skrypt załaduje się do końca. Za umowną granicę szybkości ładowania strony, powyżej której większość użytkowników będzie się czuła niekomfortowo, uznaje się 100ms. (por. np. <http://www.websiteoptimization.com/speed/tweak/psychology-web-performance/>).

72 Por. [http://www.w3schools.com/tags/tag\\_meta.asp](http://www.w3schools.com/tags/tag_meta.asp).

W sekcji head znajdują się odpowiednie tagi <meta>, zaś w sekcji body – informacje wyświetlane w tabelkach widocznych dla użytkownika.

Zazwyczaj na stronie HTML z rekordem bibliograficznym znajduje się link do pliku z tekstem artykułu, co oznacza, że te same informacje w całym układzie pojawiają się trzykrotnie, a nawet czterokrotnie: w sekcji head kodu HTML, w sekcji body kodu HTML oraz w pliku z tekstem artykułu (w przypadku pliku PDF najczęściej jeszcze w nagłówku pliku). Sytuację tę ilustruje rysunek 2.

**Rysunek 2.** Występowanie danych bibliograficznych w rekordzie: plik HTML i podlinkowany do niego plik PDF



Cyframi 1–4 zaznaczono na rysunku poszczególne miejsca, w których ta sama informacja pojawia się w różnej postaci, przy czym 1 i 2 związane są ze stroną internetową zawierającą rekord bibliograficzny, zaś 3 i 4 – z połączonym z tą stroną za pomocą linku plikiem zawierającym publikację naukową. Warto zaznaczyć, że w przypadku 1 i 3 odczytanie informacji przez automat indeksujący jest na tyle trywialne, że możemy założyć, iż najczęściej crawler dokonuje w tych miejscach wyłącznie prostego parsingu<sup>73</sup>. Odwrotnie zaś w przypadku 2 i 4 – tutaj najprawdopodobniej do odczytu danych trzeba będzie użyć bardziej skomplikowanych narzędzi do ekstrakcji.

<sup>73</sup> Jedynym wyjątkiem może być tutaj sytuacja, w której w jednym tagu <meta> zawarto całe cytowanie bibliograficzne wg jednego z wielu standardów zapisu cytowania (mowa tu o takich standardach jak MLA, APA czy ISO 690, czyli standardach normalizujących opracowanie danych możliwych do odczytania przez człowieka). Program nie ma łatwego sposobu na to, by dowiedzieć się, w jakim standardzie jest skonstruowane cytowanie, które aktualnie odczytuje. Jest to zresztą jeden z powodów, dla których rekomendujemy rozbięcie metadanych o cytowaniu bibliograficznym na wiele poszczególnych tagów <meta>, z których każdy odpowiedzialny jest za jednostkę informacji – lub przynajmniej podanie cytowania bibliograficznego w formacie przystosowanym do maszynowego parsowania. Patrz sekcje 5.5.3.1. i 5.5.3.4.

Warto też zauważyć, że żadne zewnętrzne ograniczenia nie wymuszają zgodności informacji zawartych w 1–4. To, że dane zawarte zazwyczaj w tych miejscach są egzemplarzami tej samej informacji podanej w różny sposób, jest wyłącznie użyteczną konwencją. System ten jest więc teoretycznie podatny na manipulacje. Praktyki takie mogą pojawić się przy próbie fałszowania automatycznie generowanych statystyk dotyczących danego naukowca, czasopisma lub instytucji. Google, co zrozumiało, chce się zabezpieczyć przed możliwymi sztuczkami, np. wystawieniem dokumentu, którego metadane nie będą zgodne z jego zawartością. Analogiczne praktyki spotykane były bardzo często w dziedzinie SEO dla biznesowych stron internetowych. Korzyścią dla pracowników naukowych z tego rodzaju oszustwa byłby np. możliwy do pokazania w CV indeks Hirscha wyliczony dla konta na Google Scholar. Naukowcy z EC3 Research Group pokazali, że oszukanie systemu Google Scholar nie jest trudne – dzięki maszynowemu wygenerowaniu dużej ilości odpowiednio spreparowanych, fałszywych publikacji naukowych autorzy zdołali wywindować indeks Hirscha rzekomego właściciela – założonego specjalnie w tym celu – fałszywego konta w Google Scholar<sup>74</sup>.

Porównanie metadanych wstawionych w HTML czy w nagłówku PDF z danymi wyekstrahowanymi z tekstu publikacji pozwoli w większości wypadków wykryć próby oszustwa (czy ewidentne pomyłki w metadanych). Ekstrakcja taka jest więc raczej środkiem bezpieczeństwa niż źródłem dobrych jakościowo metadanych. O tym, dlaczego wciąż jeszcze ekstrakcja nie może zastąpić wprowadzanych „ręcznie” metadanych – patrz sekcja 5.4.3.

Ponadto warto zauważyć, że rys. 2 przedstawia rozwiązanie modelowe. Rzeczywiste strony internetowe odbiegać mogą od niego na różne sposoby, z których najczęściej spotykane są dwa:

- od strony z rekordem bibliograficznym nie prowadzi link do publikacji naukowej (rekord bez pełnego tekstu) – sytuacja ta najczęstsza jest w bazach bibliograficznych i niektórych repozytoriach;
- strona zawierająca publikację lub link do niej nie zawiera metadanych bibliograficznych (pełny tekst bez rekordu) – sytuacja ta najczęstsza jest w przypadku stron domowych pracowników naukowych.

Automaty indeksujące przygotowane są na maksymalnie wiele ewentualnych odchyień od normy i np. w sytuacji braku metadanych na stronie internetowej próbują je odzyskać z nagłówka pliku PDF lub z jego warstwy tekstowej. Pozorna redundancja informacji spełnia też ważną rolę w kontroli jakości i prawdziwości danych wprowadzonych w każde z czterech wyróżnionych miejsc.

W tym miejscu może pojawić się pytanie: skoro automaty indeksujące teoretycznie mogą wyekstrahować metadane z nagłówka pliku PDF, a nawet z samego tekstu publikacji naukowej, to czy warto w ogóle inwestować w opracowanie metadanych? By na nie odpowiedzieć, musimy przynajmniej ogólnie zarysować obiektywne trudności związane z ekstrakcją metadanych z materiału przygotowanego do odczytania przez człowieka. Przegląd wyników wybranych badań pozwoli nam też zdać sobie sprawę z dokładności wyników, na jaką możemy liczyć w przypadku ekstrakcji metadanych.

---

74 Por. E.D. Lopez-Cozar, N. Robinson-García, D. Torres-Salinas, *Manipulating Google Scholar citations and Google Scholar metrics: Simple, easy and tempting*. arXiv preprint arXiv:1212.0638 (2012).

### 5.4.3. Ekstrakcja metadanych

Maszynowa ekstrakcja metadanych nie jest zadaniem prostym obliczeniowo. Nawet w tak – zdawałoby się – prostym przypadku, jak wydobycie metadanych dotyczących autora, tytułu publikacji, rocznika, tytułu czasopisma i nazwy wydawcy.

Przykładowo, metoda ekstrakcji metadanych opisana w artykule dotyczącym systemu CERMINE ma w przypadku parsowania danych z cytowania bibliograficznego skuteczność średnią rzędu 93%, zaś w przypadku ekstrakcji metadanych z całości artykułu – 77,5%. Jeśli chodzi o pojedyncze metadane, statystyki wyglądają nieco lepiej. Dla tytułu, czyli jednej z najważniejszych danych, skuteczność ekstrakcji wynosi 97,8%, zaś dla nazwiska autora – 96,5%<sup>75</sup>.

Warto jednak zdać sobie sprawę z tego, jak skomplikowane algorytmy potrzebne są do osiągnięcia takich rezultatów. CERMINE najpierw określa pozycję i rozmiar każdego znaku w artykule po to, by następnie na tej podstawie (obliczając dla każdego znaku w tekście odległości w pionie i w poziomie od pięciu najbliższych sąsiednich znaków<sup>76</sup>) wyznaczyć w tekście linie, strefy różnego przeznaczenia oraz kolumny. Te geometryczne właściwości tekstu pełnią w procesie ekstrakcji ważną rolę. Dopiero na podstawie takiego „wstępnego rozpoznania” program ustala, gdzie może szukać cytowań bibliograficznych, a skąd ekstrahować metadane dotyczące samego artykułu. Dopiero identyfikacja znaków i segmentacja stron pozwala na dokonanie właściwych ekstrakcji: metadanych oraz informacji związanych z cytowaniami<sup>77</sup>. Praca *Citation parsing using maximum entropy and repairs*<sup>78</sup>, w której prezentowane są podstawowe założenia systemu ParsCit<sup>79</sup>, porównuje jego wyniki m.in. z systemami opartymi o ukryte modele Markova. Zdefiniowane w niej zostaje robocze pojęcie „dobrze sparsowanego cytowania” (ang. *good parse*) jako cytowania, w którym co najwyżej jedno pole zostało sparsowane źle. Okazuje się, że ostatecznie (oba systemy – HMM i ParsCit – posiadały możliwość automatycznego poprawiania błędów) ParsCit „dobrze sparsował” 88% cytowań, HMM zaś – 31%<sup>80</sup>. ParsCit wykazał się w testach skutecznością 98,7% dla pola „autor”, 90,9% dla pola „tytuł” i 83,4% dla pola „miejsce wydania”<sup>81</sup>. Błędy w średnio 10% pojedynczych informacji ze zbioru tych, które przedtem uznaliśmy za najważniejsze, i 12% całych rekordów bibliograficznych z więcej niż jednym błędem – to wciąż bardzo dużo. Zasoby, jakie

---

75 D. Tkaczyk et al., *CERMINE: automatic extraction of structured metadata from scientific literature*, „International Journal on Document Analysis and Recognition (IJ DAR)”, 2015, s. 1–19. Warto porównać te wyniki z opublikowanymi w o trzy lata wcześniejszej pracy D. Tkaczyk et al., *A modular metadata extraction system for born-digital articles*, „Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on. IEEE”, 2012.

76 D. Tkaczyk i in. *CERMINE...*, s. 4.

77 Ibidem.

78 N.Y. Kiat, *Citation parsing using maximum entropy and repairs*, Tech. rep., National University of Singapore, 2005

79 System ParsCit wykorzystuje algorytmy oparte o modelowanie maksymalnej entropii. Por. N.Y. Kiat, op. cit., s. 8: 'Maximum entropy (ME) is a probability distribution modeling technique. The principle of ME modeling is to model all that is known and assume nothing about which is unknown. In other words, given a training dataset, choose a model consistent with all the constraints, but otherwise as uniform as possible.' („Maksymalna entropia (ME) jest techniką modelującą dystrybucję prawdopodobieństwa. Główną zasadą modelowania ME jest modelowanie wszystkiego, co znane przy jednoczesnym niezakładaniu niczego na temat tego, co nieznanie. Innymi słowy: dla danego treningowego zbioru danych należy wybrać model niesprzeczny ze wszystkimi zadanymi ograniczeniami, a jednocześnie na tyle jednolity, na ile to możliwe” – tłum. T. Lewandowski).

80 Op. cit., s. 29.

81 Op. cit., s. 28.

trzeba by było poświęcić na tego typu zadania, też są nie do pominięcia. Google najwyraźniej uważa je za zbyt trudne do przeprowadzenia: Google Scholar prowadzi jedynie ekstrakcję podstawowych danych z pliku PDF dla porównania z metadanymi wystawionymi przez repozytorium. Warto zwrócić uwagę, że tam, gdzie serwis nie wystawia dobrych metadanych w HTML – jak np. w przypadku ResearchGate czy (najczęściej) stron domowych naukowców – Google Scholar rzadko kiedy wychodzi poza ich podstawowy zestaw. A zatem przynajmniej w najbliższej przyszłości automatyczna ekstrakcja metadanych nie zastąpi ich dobrego opracowania.

Powyższy – z konieczności skrótowy – przegląd zagadnień związanych z repozytoriami i widocznością ich zasobów w Internecie stanowi tło dla rekomendacji, które przedstawimy w następnej części niniejszego rozdziału.

## 5.5. Rekomendacje dla repozytoriów

### 5.5.1. Wybór oprogramowania

Podstawowym wyborem stojącym przed instytucją zamierzającą założyć własne repozytorium jest decyzja dotycząca oprogramowania. Zalecamy użycie oprogramowania open source. Do najpopularniejszych należą: DSpace, EPrints oraz Digital Commons. Wybór wolnego lub otwartego oprogramowania skutkuje radykalnym ograniczeniem nakładu pracy i kosztów związanych z uruchomieniem repozytorium, a także – jak można wnioskować z rekomendacji Google Scholar – najprawdopodobniej wpłynie pozytywnie na widoczność zasobów w indeksach wyszukiwarek akademickich<sup>82</sup>. Zdecydowana większość problemów, jakie stoją przed twórcami oprogramowania repozytoryjnego, została już rozwiązana – wybierając istniejące rozwiązanie unika się wyważania otwartych drzwi. Bardzo istotne jest również wsparcie społeczności<sup>83</sup> dla najpopularniejszych systemów. W Polsce najpopularniejszym aktualnie oprogramowaniem repozytoryjnym jest DSpace (korzysta z niego 11 repozytoriów: AMUR, Repozytorium Uniwersytetu Łódzkiego, Repozytorium Uniwersytetu Kazimierza Wielkiego, Repozytorium Polsko-Japońskiej Akademii Technik Komputerowych, Repozytorium Uniwersytetu Białostockiego, eRIKA, RUM@K, Repozytorium WSB-NLU, Repozytorium Uniwersytetu Warszawskiego, Repozytorium Biblioteki Narodowej oraz Repozytorium CeON). Drugim pod względem popularności systemem jest dLibra (10 repozytoriów). Na dLibrze działa jednak również większość bibliotek cyfrowych w Polsce, więc gdyby liczyć biblioteki cyfrowe razem z repozytoriami, to dLibra byłaby najpopularniejszym oprogramowaniem.

---

82 „If you're a university repository, we recommend that you use the latest version of Eprints (eprints.org), Digital Commons (digitalcommons.bepress.com), or DSpace (dspace.org) software to host your papers. If you use a less common hosting product or service, or an older version of these, please read this entire documentation and make sure that your website meets our technical guidelines” (Jeśli reprezentujesz repozytorium uniwersyteckie, rekomendujemy, byś w celu udostępniania artykułów swojej uczelni używał najnowszej wersji oprogramowania Eprints, Digital Commons lub DSpace. Jeśli używasz mniej popularnego oprogramowania lub starszych wersji wymienionych systemów, przeczytaj całą dostępną tu dokumentację i upewnij się, że Twoja strona spełnia wszystkie nasze wymagania techniczne – tłum. T. Lewandowski) – <https://scholar.google.com/intl/en/scholar/inclusion.html>.

83 Por. <http://dspace.2283337.n4.nabble.com/>.

## 5.5.2. Widoczność zasobów repozytorium w wyszukiwarkach internetowych

Widoczność w wyszukiwarkach internetowych jest kluczowa, jeśli chodzi o widoczność zasobów repozytoryjnych w ogóle. Celem optymalizacji repozytorium pod tym względem winna być przede wszystkim poprawa widoczności w wyspecjalizowanych wyszukiwarkach akademickich, jednak wyszukiwarki ogólnego przeznaczenia nie powinny zostać przy tym całkowicie pominięte. Po pierwsze, część użytkowników znajduje zasoby repozytoryjne głównie za pośrednictwem wyszukiwarek ogólnych. Po drugie, z powodu podobnych zasad działania obu typów wyszukiwarek oraz częściowo wspólnych procesów ewaluacji stron, niektóre działania mające na celu poprawienie rankingu w wyszukiwarkach ogólnego przeznaczenia zaowocują również zwiększeniem widoczności w wyszukiwarkach akademickich. Poniżej zajmiemy się głównie optymalizacją pod kątem specjalistycznych, akademickich wyszukiwarek. Tam, gdzie rekomendacja będzie miała ogólniejszy charakter, zaznaczymy to.

### 5.5.2.1. Mapa strony

Jedną z ważniejszych praktyk SEO jest pozostawienie dla robotów indeksujących „mapy strony” (*sitemap*) – pliku XML, którego zawartość pozwoli robotowi indeksującemu łatwiej nawigować wśród poszczególnych podstron.

Mapy stron zostały wprowadzone przez Google w 2005 roku, a inne wyszukiwarki dość szybko podchwyciły pomysł; powstał odpowiedni protokół map stron (*sitemap protocol*), który od tego czasu zasadniczo się nie zmienia. Od 2011 roku Google zaczęło ignorować sitemapy umieszczone w miejscach innych niż strona główna. Wymusiło to przestrzeganie konwencji, co do tamtej chwili było jedynie zalecane.

Prawdopodobnie najważniejszą cechą mapy strony z punktu widzenia repozytorium jest możliwość nadania poszczególnym podstronom odpowiedniego priorytetu; pozwala to na dodatkowe podkreślenie podstron zawierających zdeponowane zasoby (w odróżnieniu od stron wyszukiwania czy innych – „nieistotnych” z perspektywy robotów indeksujących – rejonów repozytorium).

Jeśli repozytorium zostało utworzone niedawno i nie jest jeszcze obecne w wyszukiwarkach internetowych, możliwość przesłania mapy strony do danej wyszukiwarki poprzez narzędzia dla webmasterów udostępniane przez producenta wyszukiwarki (Google Webmaster Tools dla Google/Google Scholar, Microsoft Webmaster Tools dla Bing/Microsoft Academic Search itp.) jest dobrym sposobem na uwzględnienie przez roboty indeksujące. Gdyby nie ta możliwość, zaistnienie w wyszukiwarkach byłoby problematyczne: robot indeksujący może trafić na nową stronę poprzez linki umieszczone na stronach już istniejących, ale niewiele osób dowiaduje się o istnieniu stron, których jeszcze nie ma w indeksach wyszukiwarek, co w rezultacie sprawia, że mało osób umieszcza do nich linki. Plik *sitemap* dla nawet średniej wielkości repozytorium będzie stosunkowo duży, istnieje jednak szereg możliwości automatycznego wygenerowania mapy strony, również z pomocą narzędzi dostępnych online.

Należy również zwrócić szczególną uwagę na to, by zawartość mapy strony nie stała w sprzeczności z zawartością pliku *robots.txt* (patrz: następny podrozdział). Roboty indeksujące są wrażliwe na takie konflikty, występujące wtedy, gdy np. *sitemap* „zaprasza” pajączka do odwiedzin podstron, których

ostatecznie robots.txt nie pozwala mu zaindeksować. Możemy być pewni, że podobna strata zasobów przeznaczonych na indeksowanie zostanie odnotowana. Podobne niezgodności mogą z łatwością powstać, gdy za mapę strony i plik robots.txt odpowiadają różne osoby<sup>84</sup> lub jeśli powstały one w dłuższym odstępie czasowym.

#### 5.5.2.2. Plik robots.txt

Zarówno eksperci z Google Scholar<sup>85</sup>, jak i K. Arlitsch oraz P. O'Brien – autorzy poradnika o SEO dla repozytoriów – mocno podkreślają względnie łatwą do usunięcia, ale często spotykaną przeszkodę w indeksowaniu repozytoriów, jaką jest niewłaściwie skonfigurowany plik robots.txt. Ten krótki plik tekstowy, zamieszczony wśród plików dołączonych do strony internetowej, jest konwencjonalnym sposobem na instruowanie crawlerów internetowych, co powinny, a czego nie powinny indeksować spośród zawartości strony. Podkreślmy konwencjonalność tych instrukcji: pajęczki np. Google, Microsoftu czy Yahoo! będą stosować się do poleceń tam zawartych, ale w rzeczywistości nic ich do tego nie zmusza. Wszelkie „agresywne” crawlery będą więc ignorowały zawarte tam polecenia. Plik robots.txt używany jest np. do powstrzymywania robotów przed indeksowaniem stron będących wynikami wyszukiwań z pomocą wewnętrznej wyszukiwarki strony lub w sytuacji, gdy strona jest w wersji testowej i nie powinna być publicznie dostępna.

Zdarzają się jednak (na całym świecie i dość powszechnie) zadziwiające wypadki zamieszczenia na stronie repozytorium pliku robots.txt blokującego dostęp wyszukiwarkom do całej zawartości strony. W Polsce odnotowaliśmy tylko jeden taki przypadek, warto jednak pamiętać o tej kwestii przy zakładaniu nowego repozytorium. Eksperci z Google Scholar rekomendują robots.txt dopuszczający crawlery do całej zawartości repozytorium (tj. zostawienie pustego atrybutu „Disallow”)<sup>86</sup>.

#### 5.5.2.3. Przeglądanie

Crawlery, chcąc odnaleźć w repozytoriach rekordy, posługują się metodą przeglądania (*browse*). Ścieżki, którymi się poruszają, mają strukturę drzewa. Google Scholar rekomenduje układ: rok–miesiąc–publikacja, w którym po przejściu dwustopniowej ścieżki linków crawler znajdzie się na stronie zawierającej linki do wszystkich publikacji dostępnych w repozytorium wydanych w określonym miesiącu. Taka topologia pozwala z jednej strony na znalezienie wszystkich publikacji w trzech „krokach” crawlera (a więc drzewo linków nie jest zanadto wysokie), z drugiej zaś – nie pozwala zgubić się mu w zbyt wielu odgałęzionych ścieżkach wychodzących z jednego miejsca (a więc drzewo linków nie jest zanadto szerokie).

#### 5.5.2.4. Brak stron pośrednich między rekordem a pełnym tekstem publikacji

Kwestią, na którą należy zwrócić uwagę przy tworzeniu nowego repozytorium, jest taka architektura systemu repozytoryjnego, która sprawi, że po kliknięciu w link prowadzący do rekordu lub samego pliku z artykułem użytkownik zostanie przekierowany bezpośrednio do danego rekordu lub pliku, bez żadnych przekierowań do wyskakujących okienek, stron logowania lub stron z regulaminem korzystania

84 Por. K. Arlitsch, P. O'Brien, *Improving...*, op. cit., s. 18.

85 D. Dapra, *Indexing Repositories*, s. 11; A. Acharya, *Indexing Repositories...*, s. 35.

86 D. Dapra, op. cit., s. 14, A. Acharya, op. cit., s. 35.



z serwisu. Linki, które udostępniemy na stronie, będą tymi samymi linkami, które wyszukiwarka akademicka zaprezentuje użytkownikowi – pozostawienie go bez kontekstu na stronie logowania lub regulaminu sprawi, że najprawdopodobniej opuści serwis<sup>87</sup>. Co więcej, Google przyznaje, że woli taką pozycję usunąć z indeksu, niż narażać użytkowników na niepożądane doświadczenia<sup>88</sup>.

#### 5.5.2.5. Linki

PageRank danej strony – w tym strony repozytorium – zależy od liczby linków prowadzących do niej z innych stron w Internecie oraz od PageRanku tych stron. Warto zadbać o PageRank strony repozytorium w sposób bezpośredni, tj. upewniać się, że strony stojące wysoko w rankingu linkują do strony repozytorium. Korzyść, jaką w ten sposób się odnosi, jest podwójna. Przede wszystkim zmienia się ranking strony repozytorium w indeksach wyszukiwarek, a więc zwiększa się prawdopodobieństwo, że użytkownik wpisujący odpowiednie słowa kluczowe zobaczy link do naszego repozytorium. Warto pamiętać, że liczby w rankingu odzwierciedlają też pewne wartości. Jedną z nich jest poziom jakości informacji, którą użytkownik może znaleźć na stronie mającej wysoki PageRank. Dodatkową korzyścią jest więc to, że przez obecność na stronach wysoko stojących w rankingu PageRank informacja o danym repozytorium zwróci uwagę większej liczby użytkowników zainteresowanych zdeponowanymi w nim materiałami.

Adresem docelowym może być oczywiście strona domowa (np. jeśli dział promocji danej instytucji zamieści notkę informującą o otwarciu repozytorium, względnie o aktualnej liczbie zdeponowanych materiałów). Najbardziej oczywistym rozwiązaniem jest jednak kierowanie użytkowników do tego, co najistotniejsze w repozytorium – czyli do samych publikacji. Wszelkie blogi naukowe, czy to prowadzone przez instytucję macierzystą, czy po prostu przez naukowców korzystających z zasobów repozytorium, będą dobrym źródłem linków. Dwa źródła linków są jednak szczególnie cenne z dwóch różnych względów: media społecznościowe i Wikipedia.

##### 5.5.2.5.1. Linki w mediach społecznościowych

Media społecznościowe posiadają zazwyczaj największy możliwy PageRank – 10. Jest to zrozumiałe: aktualny rozkład ruchu w Internecie wygląda w ten sposób, że to właśnie media społecznościowe stanowią dla ogromnej większości użytkowników przedsięwzięcie znacznej części przeglądanych treści. Gdy w repozytorium zdeponowana zostanie praca szczególnie zasługująca na wyróżnienie, być może ktoś zwróci na nią uwagę samodzielnie – warto jednak wspomóc taką ewentualność. Jednym ze sposobów jest uświadamianie autorom, że w ich interesie leży promowanie swoich prac za pomocą mediów społecznościowych. Autorzy regularnie korzystający z tych mediów posiadają ukształtowaną sieć kontaktów, dzięki czemu wiadomość o nowej publikacji trafi zapewne do odpowiednich użytkowników (przynajmniej jeśli bierzemy pod uwagę środowisko akademickie). Takich autorów jest w Polsce wciąż niewielu, jednak coraz więcej osób komunikuje się z innymi członkami społeczności akademickiej właśnie za pośrednictwem serwisów społecznościowych.

---

87 Por. *Jak poprawić widoczność rekordów w Google Scholar*, op. cit.

88 Por. D. Dapra, op. cit., s. 17–20.

Drugim sposobem zadbania o widoczność zasobów repozytoryjnych w mediach społecznościowych jest prowadzenie oddzielnego konta repozytorium lub zadbanie o to, by konto reprezentujące instytucję macierzystą odpowiednio często uwzględniało repozytorium w swej działalności.

#### 5.5.2.5.2. Linki w Wikipedii

Drugim miejscem w Internecie, które szczególnie dobrze nadaje się do promocji zasobów zdeponowanych w repozytoriach, jest Wikipedia. Ta największa internetowa encyklopedia musiała przebyć długą drogę, by zyskać zaufanie pracowników naukowych. W przeciwieństwie do wielu innych internetowych źródeł wiedzy, Wikipedia może wykazać się sprawnie działającym systemem redakcji zamieszczanych w niej treści.

Wikipedia jest często jednym z pierwszych miejsc, w którym użytkownicy rozpoczynają zbieranie informacji na interesujący ich temat. Użytkownicy Internetu otwierają około 8,5 miliona stron Wikipedii w ciągu jednej godziny<sup>89</sup>. Dla polskiej sekcji Wikipedii odpowiednia liczba równa jest 248 tysiącom<sup>90</sup>. Warto w tym kontekście zwrócić uwagę na źródła bibliograficzne podane przy artykułach. W odnośnikach tych tkwi wielki potencjał promocji repozytoriów. Czytelnicy Wikipedii są dobrą grupą docelową dla artykułów naukowych. Zarazem Wikipedia ma i – jeśli wierzyć pracownikom Google<sup>91</sup> – będzie miała wysoki PageRank. Rola Wikipedii w promocji treści naukowych jest coraz powszechniej uznawana; sugestia linkowania w Wikipedii do artykułów zdeponowanych w repozytoriach znalazła się również w rekomendacjach Digital Commons<sup>92</sup>. Zadanie to najprawdopodobniej będzie zresztą wykonywane po części przez autorów zdeponowanych w repozytorium prac lub po prostu przez „zwykłych” użytkowników Wikipedii. Ma ona wbudowany system wyszukiwania zewnętrznych linków z określonej domeny<sup>93</sup>, dzięki któremu można w każdej chwili zorientować się, jak wiele i jakiego rodzaju artykułów z danego repozytorium obecnych jest na Wikipedii<sup>94</sup>.

#### 5.5.2.6. Pełne teksty artykułów dostępne lokalnie

Mogłoby się wydawać, że udostępnianie użytkownikom samego linku do dostępnego gdzie indziej pełnego tekstu artykułu jest dostatecznie dobrym rozwiązaniem. Z punktu widzenia użytkownika różnica między tekstem dostępnym lokalnie (czyli w obrębie tej samej domeny, w której znajduje się strona repozytorium) a dostępnym zewnątrz może być niedostrzegalna – zwłaszcza jeśli repozytorium linkuje bezpośrednio do plików PDF. Należy jednak pamiętać, że w ten sposób traci się kontrolę nad pełnym tekstem – jego dostępność w tej sytuacji zależy wyłącznie od dostępności tekstu w domenie docelowej<sup>95</sup>. Ponadto dobra

---

89 M. Teplitskiy, G. Lu, E. Duede, *Amplifying the Impact of Open Access: Wikipedia and the Diffusion of Science*, arXiv preprint arXiv:1506.07608 (2015), s. 3. Autorzy powołują się na stronę <http://stats.wikimedia.org/EN/TablesPageViewsMonthly.htm>. Aktualna liczba (w listopadzie 2015 r.) wynosi 12,3 milionów wyświetleń stron na godzinę.

90 Por. wspomnianą powyżej stronę <http://stats.wikimedia.org/EN/TablesPageViewsMonthly.htm>.

91 Por. D. Mituzas, *Re: How did Wikipedia manage to get such a high Google PageRank?* <https://www.quora.com/How-did-Wikipedia-manage-to-get-such-a-high-Google-PageRank>.

92 Bepress, *Search Engine Optimization: How to Attract More Visitors to Your Repository, Digital Commons Reference Material and User Guides*. Paper 35, 2012, s. 2.

93 Pod adresem: <https://en.wikipedia.org/wiki/Special:LinkSearch>

94 Kilka szczegółowych rad odnośnie do prowadzenia działalności na Wikipedii znaleźć można również w K. Arslitsch, P.S. O'Brien, *Improving...*, op.cit..

95 A. Acharya, *Indexing Repositories...*, op. cit., s. 13.

widoczność tekstu nie zależy już od wysiłków repozytorium. Repozytorium zazwyczaj nie może odpowiadać np. za prawidłową konfigurację robots.txt strony w domenie zewnętrznej. Metadane dostępne dla crawlerów w formie tagów HTML mogą z przyczyn technicznych zostać powiązane wyłącznie z tekstem dostępnym lokalnie<sup>96</sup>. Ponadto, jeśli tekst w domenie zewnętrznej zostanie mimo wszystko zaindeksowany, w wynikach zaprezentowanych użytkownikowi pojawi się nazwa tej domeny, a nie danego repozytorium. Choćby więc metadane pochodziły z repozytorium i były efektem poważnego nakładu pracy jego redaktorów, użytkownik nie dowie się o tym.

### 5.5.3. Metadane

Istnieje wiele rekomendacji Google Scholar dotyczących metadanych. Zalecenia te powinny zostać przyswojone przez repozytoria szczególnie uważnie. Przede wszystkim należy pamiętać, że Google Scholar (zaprojektowane nie tylko z myślą o repozytoriach, lecz również o stronach autorów itd.), oprócz metadanych pobieranych ze strony zawierającej link do pliku PDF z artykułem, opracowuje również automatycznie zestaw porównawczy własnych metadanych na podstawie samego pliku PDF (na tej podstawie w znacznej mierze odróżnia się też artykuły naukowe od nienaukowych). W przypadku daleko idących niezgodności pomiędzy dwoma zestawami metadanych artykuł nie zostanie zindeksowany. Jest to dość oczywisty sposób zabezpieczenia się przed oszustwami w przypadku, gdyby metadane były intencjonalnie niezgodne z zawartością artykułu (np. gdyby jako artykuły naukowe prezentowane były inne treści). Zalecamy również zwrócenie uwagi na zalecenia OpenAIRE dotyczące stosowania protokołu OAI-PMH oraz formatu Dublin Core zawarte w [OpenAIRE Guidelines](#).

#### 5.5.3.1. Format metadanych

Tagi HTML zawarte w sekcji `<head>` źródła strony są głównym źródłem metadanych dla crawlerów. Z tego powodu duże znaczenie ma format, w jakim są one zapisane. Google Scholar poleca cztery formaty: Highwire Press, Eprints, BE Press oraz PRISM<sup>97</sup>. Zwróćmy uwagę, że w rekomendacjach nie został uwzględniony format Dublin Core. Co więcej, na stronie Google Scholar jest on odradzany, zaś przez Darcy Dapra z Google Scholar jest wręcz nazywany pułapką<sup>98</sup>. Konsekwentnie powtarzaniem powodem jest niedostateczna precyzja tagów uwzględnionych w tym formacie, zwłaszcza jeśli chodzi o składowe cytowania. Dublin Core jest jednak szeroko wykorzystywany w repozytoriach i ich sieciach, pozostaje też domyślnym formatem dla OpenAIRE. Zwróćmy przy okazji uwagę na fakt, że Google zrezygnował z obsługi protokołu OAI-PMH w 2008 roku (to zagadnienie szerzej omawiamy poniżej).

Biorąc to pod uwagę, rekomendujemy równoległe użycie dwóch standardów tagów HTML, w których repozytorium umieści metadane zdeponowanego zasobu: Dublin Core oraz którykolwiek z czterech wymienionych powyżej, rekomendowanych przez Google Scholar. Aktualnie w Polsce najpopularniejszym z nich jest Highwire Press. Wszystkie dane zawarte w jednym ze standardów powinny zostać zduplikowane

---

96 A. Acharya, op. cit., s. 13.

97 Por. section 2: Configuring the meta-tags: 'Google Scholar supports Highwire Press tags (e.g., citation\_title), Eprints tags (e.g., eprints.title), BE Press tags (e.g., bepress\_citation\_title), and PRISM tags (e.g., prism.title)'.  
<https://scholar.google.com/intl/en/scholar/inclusion.html#indexing>

98 D. Dapra, *Indexing Repositories...*, s. 32.

w drugim z nich. Obecność jednego formatu nie wpływa negatywnie na możliwość odczytania drugiego, zaś ilość danych przesyłanych dodatkowo w kilku czy kilkunastu liniijkach HTML nie stanowi problemu.

### 5.5.3.2. Metadane dotyczące autorów a nazwiska widoczne w artykule

Metadane dotyczące autorów powinny zgadzać się z nazwiskami widocznymi w artykule; oba zbiory nazwisk powinny być sobie równe<sup>99</sup>. W szczególności powinno się wystrzegać błędu polegającego na podawaniu w tagach HTML jako autorów osób, które w rzeczywistości autorami nie są (np. promotorów prac dyplomowych) lub nawet instytucji, w których powstały dane publikacje (rys. 3–6)<sup>100</sup>.

Rysunek 3. Przykład błędnego użycia pola DC.creator – rekord na stronie repozytorium

**Opis publikacji**

---

## Od "naukowej" wiedzy o literaturze do "świata literackości"

**Tytuł:**  
Od "naukowej" wiedzy o literaturze do "świata literackości"

**Inny tytuł:**  
Teksty Drugie Nr 5-6 (1990)

**Twórca:**  
Instytut Badań Literackich

**Wydawca:**  
IBL PAN

**Miejsce wydania:**  
Warszawa

**Data wydania:**  
1990

**Opis:**  
21 cm

**Temat i słowa kluczowe:**  
literatura ; badania literackie ; nauka

**Publikacja planowana**



Powinno się dopilnować, by wszyscy autorzy pracy byli wymienieni<sup>101</sup>. Innym często spotykanym błędem jest podawanie autorów w innej kolejności niż ta podana w publikacji<sup>102</sup>. Za kolejność wymieniania autorów publikacji odpowiedzialna jest wypracowana w środowiskach akademickich konwencja, zdarza się natomiast, że system repozytoryjny podaje ich w tagach HTML w kolejności np. alfabetycznej. W standardowym modelu działania repozytorium, w którym metadane wprowadzane są przez deponujących artykuły naukowców, należy pod tym względem kontrolować każdy wprowadzony rekord. Pomocne będzie też zamieszczenie odpowiednich instrukcji dla autorów na stronie internetowej repozytorium.

99 A. Acharya, *Indexing Repositories...*, s. 25, D. Dapra, *Indexing Repositories...*, s. 26–27.

100 Por. D. Dapra, op. cit., s. 28–29.

101 A. Acharya, op. cit., s. 22, D. Dapra, op. cit., s. 23.

102 A. Acharya, op. cit., s. 22, D. Dapra, op. cit., ss. 24–25.

Rysunek 4. Przykład błędnego użycia pola DC.creator – tagi <meta> w sekcji head pliku HTML

```
<meta name="DC.relation" content="27. S. J. Schmidt, Fictionality in Literary and non-Literary Discourse, "Poetics" 1980, nr 5-6.">
<meta name="DC.relation" content="28. S. J. Schmidt, O pisaniu historii literatury. Kilka uwag ze stanowiska konstruktywistycznego, przeł. M. B. Fedewicz, "Pamiętnik Literacki" 1988, z. 3.">
<meta name="DC.relation" content="29. T. Sławek, Via femina. Dekonstrukcja jako styl krytyki, w: Interpretacje i style krytyki, red. W. Kalaga, T. Sławek, Katowice 1988, s. 149.">
<meta name="DC.relation" content="30. J. Sławiński, Socjologia literatury i poetyka historyczna, w: idem, Dzieło, język, tradycja, Warszawa 1974, s. 65-67.">
<meta name="DC.relation" content="31. J. Sławiński, Związki metodologiczne, "Teksty" 1978, z. 5.">
<meta name="DC.relation" content="Teksty Drugie">
<meta name="DC.identifier" content="0867-0633">
<meta name="DC.creator" content="Instytut Badań Literackich">
<meta name="citation_publication_date" content="1990">
<meta name="citation_lastpage" content="32">
<meta name="citation_keywords" content="literatura">
<meta name="citation_keywords" content="badania literackie">
<meta name="citation_keywords" content="nauka">
<meta name="citation_title" content="Od "naukowej" wiedzy o literaturze do "świata literackości">
<meta name="citation_doi" content="0867-0633">
<meta name="citation_firstpage" content="16">
<meta name="citation_authors" content="Instytut Badań Literackich">
<meta name="citation_author" content="Instytut Badań Literackich">
<meta name="citation_issue" content="5-6">
<meta name="citation_journal_title" content="Teksty Drugie">
<meta name="citation_date" content="1990">
<meta name="citation_language" content="pol">
<meta name="citation_publisher" content="IBL PAN">
<style type="text/css">...</style>
</head>
<body>
```

Jeden z polecanych przez Google Scholar formatów, Highwire Press, uwzględnia dwa rodzaje tagów dla autorów: citation\_authors (liczba mnoga) i citation\_author (liczba pojedyncza). Wszystkie przykłady dobrych praktyk prezentowane przez pracowników Google Scholar zawierają szereg tagów „pojedynczych” (citation\_author) – należy więc zakładać, że jest to opcja optymalna.

Rysunek 5. Przykład skutków błędnego użycia pola DC.creator – strona z wynikami w Google Scholar



Kwestia właściwej kolejności autorów w metadanych staje się szczególnie ważna, jeśli zdecydujemy się na użycie tagów pojedynczych (citation\_author). Należy upewnić się, że system repozytoryjny wygeneruje tagi „citation\_author” w źródle strony w kolejności ich podania, a nie w kolejności alfabetycznej lub pseudolosowej.

**Rysunek 6.** Przykład skutków błędnego użycia pola DC.creator – błędne cytowanie przygotowane automatycznie przez Google Scholar

**Cytuj**

Skopiuj i wklej sformatowany opis bibliograficzny lub użyj jednego z linków, aby importować dane do menedżera bibliografii.

**MLA** Literackich, Instytut Badań. "Od" naukowej" wiedzy o literaturze do" świata literackości". *Teksty Drugie* 5-6 (1990): 16-32.

**APA** Literackich, I. B. (1990). Od" naukowej" wiedzy o literaturze do" świata literackości". *Teksty Drugie*, (5-6), 16-32.

**ISO 690** LITERACKICH, Instytut Badań. Od" naukowej" wiedzy o literaturze do" świata literackości". *Teksty Drugie*, 1990, 5-6: 16-32.

[BibTeX](#) [EndNote](#) [RefMan](#) [RefWorks](#)

W treści taga powinno pojawić się wyłącznie imię i nazwisko autora. Na dalsze dane, takie jak afiliacja czy stopień lub tytuł naukowy, przewidziano odpowiednie tagi dodatkowe. Należy pamiętać, że niektóre tagi służą do automatycznego wygenerowania cytowania. Nie powinno się w nich umieszczać niczego, czego by się nie umieściło w normalnym cytowaniu bibliograficznym.

### 5.5.3.3. Metadane dotyczące daty

Metadane dotyczące daty powinny odnosić się jedynie do daty wydania artykułu i zawierać tylko rok<sup>103</sup>. Trzeba pamiętać, że data zostanie użyta w konstrukcji cytowania, z której skorzystać będzie mógł użytkownik końcowy<sup>104</sup>. Data niezgodna z formatami uniemożliwi mu łatwe skorzystanie z tej opcji. Google Scholar zwraca uwagę na częste przypadki wpisywania daty digitalizacji czy zamieszczenia pozycji w repozytorium. Popelniony tutaj błąd może być szczególnie ważny dla widoczności artykułu, gdyż niezgodność daty rocznej z informacjami zawartymi w pliku PDF lub z metadanymi umieszczonymi przy innych kopiach tego samego artykułu (np. w innym repozytorium) może spowodować wykreślenie rekordu z indeksu Google Scholar.

### 5.5.3.4. Cytowania

Ponieważ metadane służą akademickim wyszukiwarkom do automatycznej konstrukcji cytowania, należy zadbać o to, by znalazły się wśród nich również potrzebne do tego informacje. Oprócz powszechnie występujących w polskich repozytoriach informacji o autorze, tytule i roczniku publikacji niezbędne są również dane o tytule, tomie i numerze czasopisma oraz numeracji stron.

Google Scholar wielokrotnie<sup>105</sup> podkreślało, że używanie wyłącznie standardu Dublin Core może

103 A. Acharya, op. cit., s. 23.

104 D. Dapra, op. cit., s. 31.

105 A. Acharya, op. cit., s. 21, D. Dapra, op. cit., s. , a także Google, *Google Scholar Inclusion Guidelines*, section Indexing: „Use Dublin Core tags (e.g., DC.title) as a last resort - they work poorly for journal papers because Dublin Core doesn't have unambiguous fields for journal title, volume, issue, and page numbers”. (Używaj tagów Dublin Core (tj. np. DC.title)

skutkować zmniejszoną widocznością rekordów w ich indeksie. Konsekwentnie podaje się dwa powody. Pierwszym jest brak odpowiednich pól dla kluczowych w przypadku artykułów publikowanych w czasopiśmie naukowych informacji składających się na cytowanie: tomu, zeszytu, numeru i numeracji stron. Drugim – niejednolite użycie niektórych pól standardu Dublin Core wśród wydawców i repozytoriów. Pola dc.creator, dc.relation i dc.source mają być tu źródłem szczególnie częstych przypadków niejednoznaczności<sup>106</sup>. Warto może zaznaczyć, że gdy pracownicy Google Scholar mówią o braku pól potrzebnych do uzyskania danych składających się na cytowanie, mają na myśli raczej praktykę niż teorię. Standardy Dublin Core przewidują zawarcie informacji o cytowaniu w specjalnie do tego przeznaczonym polach<sup>107</sup>, w szczególności w dcTERMS.bibliographicCitation<sup>108</sup>.

Co więcej, może być ono użyte na dwa sposoby. Pierwszym i domyślnym jest po prostu zawarcie w tym polu całości cytowania (tak jak by się je np. zawarło w przypisie bibliograficznym lub bibliografii). Standardy przewidują jednak również podanie informacji o cytowaniu w taki sposób, by było ono gotowe do łatwego maszynowego odczytania. Przykładowy tag z metadanymi gotowymi do automatycznego odczytania przedstawia rysunek 7.

#### Rysunek 7. Pole dcTERMS.bibliographicCitation przystosowane do automatycznego odczytania (machine-readable)

```
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<link rel="schema.KEV" href="info:ofi/fmt:kev:mtx:" />

<meta name="DCTERMS.bibliographicCitation"
content="Library and Information Science Research 22(3), 311-338" />

<meta name="DCTERMS.bibliographicCitation" scheme="KEV.ctx"
content="&ctx_ver=Z39.88-2004
&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal
&rft.jtitle=Library+and+Information+Science+Research&rft.stitle=LISR
&rft.volume=22&rft.issue=3&rft.spage=311
&rft_id=info%3Asid%2Fmimas.ac.uk%3Azetoc" />

<link rel="DCTERMS.isPartOf" href="urn:ISSN:0740-8188" />
```

Problem w tym, że prawie nikt tego nie robi. Przegląd źródeł stron z rekordami bibliograficznymi we wszystkich polskich repozytoriach częściowo potwierdził diagnozę Google. Na 25 przejrzanych repozytoriów cztery w ogóle nie zaimplementowały Dublin Core, jedno pomimo implementacji nie posiadało żadnych metadanych na temat cytowania (ponieważ w poszczególnych plikach umieszczano całe numery czasopism), jedno zawierało dane o cytowaniu w polu dc.description, dwa – w dc.relation,

---

w ostateczności – nie działają one optymalnie dla artykułów publikowanych w czasopiśmie, ponieważ Dublin Core nie przewiduje jednoznacznych pól dla tytułu czasopisma, tomu, numeru oraz numerów stron – tłum. T. Lewandowski) – <https://scholar.google.com/intl/en/scholar/inclusion.html#indexing>.

106 <http://googlewebmastercentral.blogspot.com/2008/04/retiring-support-for-oai-pmh-in.html>, a także <http://librarytechnology.org/repository/item.pl?id=9944>

107 <http://dublincore.org/documents/dc-citation-guidelines/>

108 <http://dublincore.org/documents/dcmi-terms/#terms-bibliographicCitation>

kolejne dwa – w `dc.identifier`, w pięciu dane dotyczące cytowania były niepełne (zawierały dowolny podzbiór zbioru zawierającego następujące elementy: `dc.creator`, `dc.date`, `dc.publisher`, `dc.title`), zaś dziewięć repozytoriów zawierało dane dotyczące cytowania – zgodnie z wytycznymi – w polu `dcTERMS.bibliographicCitation`. Spośród tych dziewięciu ani jedno nie zawierało jednak danych w formacie możliwym do łatwego maszynowego odczytania; we wszystkich przypadkach zawartość pola stanowił prosty ciąg znaków tekstowych zawierający cytowanie w dowolnym formacie bibliograficznym używanym w piśmiennictwie (rys. 8). Doświadczenie Google Scholar zdaje się wskazywać na to, że również na świecie tag `dcTERMS.bibliographicCitation` jest używany w podobny sposób.

Podkreślmy, że jeśli pole `dcTERMS.bibliographicCitation` zostanie użyte w sposób uznany za domyślny, tj. zostanie w nim umieszczone cytowanie w formacie *human-readable*, jego przydatność dla crawlera będzie praktycznie zerowa (w przypadku użycia wersji *machine-readable* jest ona nieznana). Polecany przez Google Scholar format Highwire (oraz formaty doń podobne) posiada poszczególne, pojedyncze dane dotyczące cytowania w oddzielnych tagach `<meta>`. Po sparsowaniu strony HTML crawler dysponuje już wszystkimi zawartymi w nich danymi. Rozwiązanie wykorzystane w przypadku `dcTERMS.bibliographicCitation` sprawia, że po sparsowaniu strony HTML crawler dysponuje jedynie gotowym do sparsowania ciągiem znaków. Dopiero jego sparsowanie (drugie już dla danej strony) sprawi, że dane w nim zawarte będą mogły zostać zaindeksowane. Pytanie o niższą wydajność `dcTERMS.bibliographicCitation` w porównaniu do tagów Highwire, choć nietrywialne<sup>109</sup>, jest w obliczu sporadycznego używania pola `dcTERMS.bibliographicCitation` w wersji *machine-readable* raczej akademickie.

Warto w tym miejscu zauważyć, że w wyniku wspomnianego przed chwilą przeglądu polskich repozytoriów pod względem metadanych dotyczących cytowania, okazało się, że praktyka dotycząca formatów polecanych przez Google Scholar (spośród nich w zasadzie jedynym używanym jest Highwire) wcale nie wygląda lepiej niż w przypadku Dublic Core. Tylko w jednym przypadku zawarto w metadanych pola `citation_journal_title`, `citation_issue` i `citation_firstpage` (nie było jednak pola `citation_lastpage`). Poza tym jednym przypadkiem, o ile metadane w ogóle uwzględniały pola w formacie Highwire (19 na 25 przypadków), zawierały się one w zbiorze `{citation_author[s], citation_title, citation_publisher, citation_date}`. Dla porównania, prowadzona przez ICM UW platforma czasopism Biblioteka Nauki, działająca w oparciu o oprogramowanie YADDA, udostępnia metadane zawierające następujące pola Highwire:

- `citation_title`
- `citation_journal_title`
- `citation_issn`
- `citation_author`
- `citation_volume`

---

<sup>109</sup> Nawet minimalne różnice w wydajności dla poszczególnych rekordów mogą dla bazy tak wielkiej, jak ta używana przez Google Scholar, oznaczać, w wyniku kumulacji, niebagatelną różnicę w zasobach niezbędnych do zgromadzenia informacji.



- citation\_issue
- citation\_date
- citation\_firstpage
- citation\_lastpage
- citation\_publisher
- citation\_doi

Podane w ten sposób metadane są właściwie gotowe do złożenia w cytowanie. Google Scholar podaje użytkownikom takie cytowania (rys. 6) – jeśli dane cytowania będą niekompletne lub błędne, Google ryzykuje złe doświadczenia użytkownika, których z zasady stara się unikać. Dlatego w niektórych przypadkach może zrezygnować z indeksowania danej pozycji – zwłaszcza gdy (co się często zdarza) inne wersje tej samej publikacji dostępne są w innych miejscach w Internecie. Może to skutkować mniejszą widocznością repozytorium, nawet gdyby crawlery nie miały problemów z dotarciem do jego zawartości.

### Rysunek 8. Przykład użycia dcTERMS.bibliographicCitation nieprzystosowanego do automatycznego odczytania (human-readable)

```
<meta name="DC.creator" content="Romaniuk, Ryszard S.">
<meta name="DCTERMS.dateAccepted" content="2013-11-12T08:12:22Z" scheme="DCTERMS.W3CDTF">
<meta name="DCTERMS.available" content="2013-11-12T08:12:22Z" scheme="DCTERMS.W3CDTF">
<meta name="DCTERMS.issued" content="1979-01-02" scheme="DCTERMS.W3CDTF">
<meta name="DCTERMS.bibliographicCitation" content="R.S. Romaniuk, Metody badania dyspersji światłowodów wielomodowych, II Krajowe Sympozjum, Światłowody i ich Zastosowania, Jabłonna 13-15 lutego 1979, Tom I, str.27-33, PAN Warszawa 1979" xml:lang="pl">
<meta name="DC.identififier" content="ZGPW, 1169.S-88, 30.12.1878">
<meta name="DC.identififier" content="https://depot.ceon.pl/handle/123456789/2748" scheme="DCTERMS.URI">
<meta name="DCTERMS.abstract" content="Artykuł przedstawia sposoby analizy dyspersji w światłowodzie wielomodowym w zależności od metody przyjętego modelu propagacji mocy optycznej we włóknie o refrakcji skokowej i gradientowej." xml:lang="pl">
<meta name="DC.description" content="Politechnika Warszawska" xml:lang="pl">
<meta name="DC.language" content="pl" xml:lang="pl" scheme="DCTERMS.RFC1766">
<meta name="DC.publisher" content="Polska Akademia Nauk, Komitet Elektroniki i Telekomunikacji" xml:lang="pl">
<meta name="DC.relation" content="Creative Commons Uznanie autorstwa 3.0 Polska" xml:lang="pl_PL">
<meta name="DC.relation" content="info:eu-repo/semantics/openAccess">
<meta name="DC.rights" content="http://creativecommons.org/licenses/by/3.0/pl/legalcode" scheme="DCTERMS.URI">
<meta name="DC.subject" content="światłowody wielomodowe" xml:lang="pl">
<meta name="DC.subject" content="optoelektronika światłowodowa" xml:lang="pl">
<meta name="DC.subject" content="światłowody włókniste" xml:lang="pl">
<meta name="DC.subject" content="włókna optyczne" xml:lang="pl">
<meta name="DC.subject" content="dyspersja światłowodowa" xml:lang="pl">
<meta name="DC.title" content="Metody badania dyspersji światłowodów wielomodowych" xml:lang="pl">
<meta name="DC.type" content="info:eu-repo/semantics/article" xml:lang="pl">
<meta name="DC.type" content="info:eu-repo/semantics/conferenceObject" xml:lang="pl">
<meta name="DC.contributor" content="Instytut Podstaw Elektroniki Politechniki Warszawskiej" xml:lang="pl">
<meta name="DC.description" content="Ryszard Romaniuk">
<meta content="Metody badania dyspersji światłowodów wielomodowych" name="citation_title">
<meta content="https://depot.ceon.pl/handle/123456789/2748" name="citation_abstract_html_url">
```

Warto podkreślić, że wytyczne dla Dublin Core przewidują dokładnie ten sam zestaw pól metadanych podanych jako wartość atrybutu tagu dcTERMS.bibliographicCitation. To, który tag <meta> w formatach polecanych przez Google Scholar odpowiada określonemu miejscu w dcTERMS.bibliographicCitation, można prześledzić wykorzystując tabelę 1.

**Tabela 1.** Zestawienie metadanych z formatów Highwire i Dublin Core w kontekście zaleceń Google Scholar

informacja	Highwire	Dublin Core	dcTERMS. bibliographicCitation
nazwisko autora	citation_author	dc.creator	TAK
tytuł artykułu	citation_title	dc.title	TAK
słowa kluczowe	citation_keywords	dc.subject	NIE
rok wydania	citation_date	dc.issued	TAK
rok wydania	citation_date	dcTERMS.issued	TAK
tytuł konferencji	citation_conference_title	dc.ispartof	TAK
tytuł czasopisma	citation_journal_title	dc.ispartof	TAK
zeszyt czasopisma	citation_issue	dc.citation.issue	TAK
tom czasopisma	citation_volume	dc.citation.volume	TAK
numer pierwszej strony	citation_fistpage	dc.citation.spage	TAK
numer ostatniej strony	citation_lastpage	dc.citation.epgae	TAK
wydawca czasopisma	citation_publisher	dc.publisher	TAK
instytucja publikująca dysertację	citation_dissertation_ institution	dc.publisher	TAK
instytucja publikująca raport techniczny	citation_technical_report_ intitution	dc.publisher	TAK
numer ISSN	citation_issn	–	NIE
numer ISBN	citation_isbn	–	NIE

#### 5.5.4. Wybrane elementy architektury strony

O architekturze strony wspominaliśmy już wcześniej. Jest to zagadnienie stanowczo zbyt rozległe, by wyczerpać je w ramach niniejszego podrozdziału. W tym miejscu postaramy się przedstawić tylko podstawowe informacje dotyczące wybranych problemów, mogących w największym stopniu wpłynąć na widoczność strony repozytorium w Internecie.

##### 5.5.4.1. Optymalizacja wydajności strony

Szybkość odpowiedzi strony internetowej na zachowania użytkownika (w tym czas ładowania) odgrywa kluczową rolę w kształtowaniu jego wrażeń. Serwisy wyszukiwarkowe (jak niejednokrotnie zaznaczaliśmy) starają się we własnym interesie zminimalizować możliwość negatywnych doznań użytkowników, w wyniku czego strony o długim okresie ładowania otrzymują niższy ranking w indeksie. Również ograniczone zasoby serwerów wyszukiwarek odgrywają tu rolę – strony odpowiadające dłużej mają mniejszą szansę na to, by w ogóle zostać zaindeksowane<sup>110</sup>.

Dlatego rekomendujemy zoptymalizowanie wydajności strony repozytorium. Niestety, użyteczne pod tym względem narzędzie Google Page Speed Service<sup>111</sup> nie jest już dostępne. Z ogólnodostępnych, darmowych i niewymagających dodatkowych instalacji narzędzi pozostaje nadal analiza czasu ładowania i wykonywania poszczególnych komponentów strony z pomocą Google Chrome Developer Tools (zakładka Timeline). Te same funkcjonalności, co Developer Tools, posiada Firebug, dodatek do przeglądarki Firefox. Z ich pomocą można

110 Por. K. Arlitsch, P. O'Brien, *Improving...*, s. 27, ramka „Speed Matters”, a także <http://googleresearch.blogspot.com/2009/06/speed-matters.html>.

111 K. Arlitsch, P. O'Brien, op. cit., s. 61–62.

rozpocząć analizę czasu ładowania strony, polecamy jednak, by jakiegokolwiek konkretne rozwiązania w tej sprawie zostały opracowane i zaimplementowane przez odpowiednich specjalistów.

#### 5.5.4.2. Optymalizacja bazy danych repozytorium

Jak dowiedzieliśmy się z sekcji 5.3.2., strona internetowa repozytorium działa na zasadzie podobnej do systemów CMS, jak np. Wordpress czy Joomla!: gotowe do wyświetlenia podstrony z poszczególnymi rekordami nie istnieją jako gotowe pliki HTML. Istnieje za to szablon, który jest wypełniany treścią wydobywaną dla każdego rekordu z działającej na serwerze bazy danych.

Długość oczekiwania na wyświetlenie strony zależy w dużej mierze od wielkości bazy danych. Im większe repozytorium, tym wolniej działa, bo tym dłużej zajmuje pozyskanie danych z bazy. W momencie, gdy dalsza optymalizacja samej strony WWW nie będzie już przynosiła widocznych rezultatów, obsługa repozytorium może podjąć decyzję o optymalizacji współpracy z bazą danych.

Pomijając te przypadki, w których repozytorium prowadzone jest w oparciu o samodzielnie opracowane oprogramowanie, odradzamy wszelkie próby optymalizacji samej architektury bazy. Dostępne rozwiązania open source są na takim poziomie zaawansowania i zostały na tyle dobrze opracowane przez społeczność, że optymalizacja w tym względzie nie ma większych szans powodzenia.

Z drugiej strony, jeśli rzeczywiście czas reakcji serwera jest „wąskim gardłem” w interakcji użytkownika z repozytorium, można rozważyć udoskonalenie sprzętu komputerowego.

#### 5.5.4.3. Przyjazne użytkownikowi adresy URL<sup>112</sup>

Tworząc ranking indeksowanych stron, Google bierze pod uwagę własności, które w jego przekonaniu mogą wpłynąć na zadowolenie użytkownika. Wśród nich wymienił należy jedną, o którą można zadbać w szczególnie łatwy sposób.

Wszystkie adresy URL, a zwłaszcza te, pod którymi znajdują się zasoby repozytorium, powinny być „przyjazne użytkownikowi”. Przyjazny URL to adres możliwie najłatwiejszy do odczytania przez człowieka. Z jego brzmienia powinna płynąć informacja, do czego prowadzi (zawarcie w nim np. tytułu lub autora publikacji będzie tu dobrym pomysłem), nie powinien natomiast zawierać zbyt długich ciągów liczbowych etc.

#### 5.5.4.4. Komunikaty o błędach

Narzędzia dla webmasterów (Webmaster Tools) oferowane zarówno przez Google, jak i Microsoft (firmę odpowiedzialną za drugą co do wielkości wyszukiwarkę internetową na rynku – Bing – oraz za serwis internetowy Microsoft Academic Research) pozwalają prześledzić ilość i rodzaj błędów, jakie powstają podczas wizyty crawlerów danej firmy w domenie repozytorium. Autorzy książki *Improving the visibility and use of digital repositories through SEO* donoszą, że początkowo narzędzia konfrontowały ich z liczbą błędów dla domeny repozytorium Uniwersytetu w Utah wynoszącą około 100 tysięcy.

---

<sup>112</sup> Por. op. cit., s. 60–61.

Na podstawie ich analizy wypracowali oni zestaw rekomendacji, których skrót, uzupełniony o nasz komentarz, przedstawiamy poniżej.

#### 5.5.4.4.1. Własne komunikaty o błędach<sup>113</sup>

Jak potwierdza Google<sup>114</sup>, indeksowane strony dostają w rankingu dodatkowe punkty za wysyłanie użytkownikowi własnych komunikatów o błędach, zwłaszcza w przypadku błędu 404. Google uważa, że domyślne komunikaty są na tyle zniechęcające dla użytkownika, że często kończy on pobyt na stronie, jeśli otrzyma niewiele mówiący komunikat o błędzie. Jeśli natomiast stronę wysyła użytkownikowi w razie błędu odpowiednio się zaprojektuje, można mu w łatwy sposób doradzić kilka scenariuszy zachowań lub nawet osłodzić przykre doświadczenie jakimś dowcipnym obrazkiem.

#### 5.5.4.4.2. Miękkie 404

Należy jednak pamiętać, że wysyłanie własnych komunikatów o błędach może w rezultacie wywołać tzw. „miękkie 404”. Własne strona błędu jest traktowana przez system jak normalna strona, do której użytkownik został przekierowany, i – jako taka – wysyła do przeglądarki użytkownika (lub do robota indeksującego) standardowy komunikat HTTP 200, czyli „wszystko dobrze”. Sytuacja taka – w omawianym przypadku będąca efektem nieuwagi – bywa jednak niekiedy wytwarzana intencjonalnie w celu zmylenia robotów indeksujących i sztucznego zawyżenia rankingu w indeksie wyszukiwarki. Jest to więc praktyka z zakresu tzw. „czarnego SEO” (*black hat SEO*) i jako taka jest zwalczana przez Google (aktualnie systemy Google z łatwością wykrywają większość przypadków miękkiego 404). Zwalczanie to odbywa się oczywiście poprzez zaniżanie rankingu stron, na których wykryto podobne praktyki; to właśnie grozi repozytorium, które nie dopilnuje tej kwestii.

#### 5.5.4.4.3. Inne komunikaty o błędach

Innymi błędami (obok File Not Found 404) najczęściej spotykanymi podczas odwiedzania stron repozytoriów mogą być:

- Forbidden 403 – użytkownik nie ma prawa dostępu do danej strony,
- Gone 410 – strona usunięta i brak adresu do przekierowania,
- Temporarily Unavailable 503 – strona tymczasowo niedostępna.

Komunikaty te warto wysyłać w opisanej powyżej formie. Choć zamiast dwóch ostatnich błędów wystąpić mogłoby zwykłe 404, to jednak wysłanie któregoś z nich daje użytkownikowi dodatkowe informacje (np. w przypadku 503 wie on, że może spróbować za jakiś czas ponownie dostać się na ten sam adres, czego w przypadku 404 mógłby już nie zrobić). Pierwszy błąd w repozytoriach otwartych powinien oczywiście występować jak najrzadziej.

---

113 Op. cit., s. 54.

114 Por. np. <https://support.google.com/webmasters/answer/93641>.

# Rozdział 6

## Działania informacyjne i komunikacyjne

Wraz z rozwojem społeczeństwa informacyjnego i wzrostem roli Internetu wyzwania komunikacyjne stające przez instytucjami naukowymi są coraz większe. Instytucje te muszą bowiem nie tylko prowadzić badania, ale również docierać do bardzo różnych grup: własnych pracowników, obecnych i potencjalnych studentów (w przypadku instytucji prowadzących działalność dydaktyczną), ośrodków naukowych w Polsce i za granicą, firm zainteresowanych nawiązaniem współpracy czy wreszcie mediów chętnie korzystających z autorytetu ekspertów ze świata nauki. Nie można też pominąć zainteresowanych nauką obywateli, dla których renoma instytucji naukowej stanowi gwarancję jakości rozpowszechnianej przez nią wiedzy. Internet nie tylko zrewolucjonizował komunikację naukową, lecz również otworzył uczelniom i jednostkom naukowym nowe możliwości kształtowania ich wizerunku.

Instytucje naukowe są w tym zakresie coraz bardziej aktywne. Wiele z podejmowanych przez nie działań – w szczególności te, które mają na celu budowanie wizerunku poprzez popularyzację prowadzonych badań – może zyskać dzięki uwzględnieniu otwartych treści naukowych. Jednostki odpowiedzialne za komunikację, takie jak działy promocji i informacji, dzięki otwartemu dostępowi zyskują możliwość wykorzystania treści, które mogą być atrakcyjne, wartościowe i ciekawe dla różnych grup odbiorców. Wpływa to korzystnie na wizerunek instytucji – udostępnianie w sposób otwarty artykułów i książek to przede wszystkim realizacja idei dzielenia się wiedzą, która wpisuje się w społeczną rolę instytucji naukowej. Ekspozowanie osiągnięć naukowych udokumentowanych publikacjami, z którymi każdy może się zapoznać, nie tylko poprawia pozycję instytucji w wyszukiwarkach i zwiększa jej rozpoznawalność, ale również ułatwia nawiązywanie współpracy z innymi ośrodkami naukowymi oraz firmami poszukującymi partnerów w środowisku akademickim. Udostępnianie doktoratów przyczynia się z kolei do zwiększenia transparentności nadawania stopni naukowych. Zanim jednak treści naukowe w otwartym dostępie stworzą osobom odpowiedzialnym za wizerunek instytucji nowe możliwości działania, a wdrożone polityki będą mogły służyć za przykład w środowisku akademickim, wsparcie komunikacyjne będzie potrzebne przede wszystkim w celu przekonania samych autorów do otwartych modeli.

## 6.1. Wprowadzanie otwartego dostępu w instytucji

Komunikacja jest jednym z kluczowych czynników decydujących o powodzeniu procesu wprowadzania otwartych rozwiązań w instytucji. Jak pokazują badania, wiele przekonań dotyczących otwartego dostępu może powstrzymać naukowców przed udostępnianiem prac w tym modelu<sup>115</sup>. Dlatego zarówno wprowadzaniu polityki otwartości, jak i tworzeniu repozytorium instytucjonalnego powinny towarzyszyć kompleksowe działania komunikacyjne mające na celu dostarczenie rzetelnych informacji na temat korzyści płynących z otwartego dostępu oraz rozwianie obaw związanych z publikowaniem w otwartych czasopismach i samoarchiwizacją publikacji.

### 6.1.1. Polityka otwartości

Pierwszym istotnym momentem w komunikacji związanej z promowaniem otwartego dostępu w instytucji może być decyzja dotycząca przyjęcia polityki otwartości i prowadzący do niej proces. Skutki wdrożenia tej polityki będą oczywiście dotyczyły znacznie szerszej grupy niż tylko osób bezpośrednio zaangażowanych w jej wypracowanie. Władzom danej instytucji będzie zależeć nie tylko na jej zrozumieniu przez pracowników, ale również na pozytywnym odbiorze, co wymaga jasnej artykulacji przyjętych założeń i celów. Nie wystarczy zatem sama informacja o wprowadzanej zmianie. Jeszcze przed jej wprowadzeniem pracownicy powinni mieć możliwość uzyskania przy pomocy narzędzi komunikacji wewnętrznej – gazety uczelnianej, intranetu czy newslettera – wiedzy o tym, czym jest otwarty dostęp, jaka jest jego geneza, jak wpływa na obieg informacji naukowej na świecie i jak wspiera globalną społeczność akademicką w dostępie do najnowszych wyników badań.

Ogłoszenie przyjętej polityki musi cechować dbałość o to, aby pracownicy zrozumieli jej praktyczne konsekwencje, czyli swoje nowe zobowiązania i możliwe sposoby ich realizacji. Od strony praktycznej wymaga to wskazania osób odpowiedzialnych za pomoc techniczną i prawną oraz przygotowania efektywnych kanałów komunikacji z tymi osobami.

Każda zmiana w organizacji – szczególnie taka, która wiąże się z dodatkowymi zadaniami dla pracowników, co ma akurat miejsce w wypadku wdrażania polityki otwartości – jest łatwiej akceptowana, jeśli osoby, których ona dotyczy, znają jej powody i cele. Z tego względu ważne jest opracowanie zestawu argumentów, których można użyć, aby ułatwić pracownikom zmianę dotychczasowych nawyków i przełamanie ewentualnych uprzedzeń. W pierwszej kolejności powinny się wśród nich znaleźć korzyści indywidualne, takie jak lepsza widoczność publikacji w Internecie, większa szansa na cytowanie czy nawiązanie współpracy z naukowcami z innych ośrodków akademickich. Warto wskazywać jednak również korzyści dla samej instytucji, takie jak poprawa wizerunku w sieci, skuteczniejsze upowszechnianie wyników realizowanych projektów badawczych, większa transparentność czy ułatwienie zarządzania wiedzą.

<sup>115</sup> *Wdrożenie i promocja otwartego dostępu do treści naukowych i edukacyjnych. Praktyki światowe a specyfika polska. Przewidywane koszty, narzędzia, zalety i wady*, opr. zespół ICM pod kierownictwem prof. Marka Niezgódki, Warszawa 2011, s. 208–218.

W dalszej kolejności pracownicy instytucji potrzebować będą informacji, które ułatwią im realizację zobowiązań wynikających z wprowadzenia polityki otwartego dostępu. Będą to przede wszystkim informacje praktyczne: jak zdobyć wiedzę o zasadach dotyczących samoarchiwizacji publikacji przez autorów, jakie przyjął dany wydawca<sup>116</sup>; co robić, jeśli nie zgadza się on na umieszczenie artykułu w repozytorium; jak założyć konto w repozytorium i zdeponować w nim publikację; jaką licencję wybrać, udostępniając swoją pracę itd.

### **6.1.2. Repozytorium instytucjonalne**

Repozytorium instytucjonalne jest znakomitym narzędziem promocji dorobku instytucji i jej pracowników – pod warunkiem, że znajduje się w nim znaczna część tego dorobku.

Szybkość, z jaką repozytorium napełni się treściami (takimi jak artykuły z czasopism naukowych, doktoraty, raporty, książki czy materiały konferencyjne) w znacznym stopniu zależy od przyjętej przez instytucję polityki otwartego dostępu. W wypadku przyjęcia otwartego mandatu można oczekiwać szybkiego tempa rozwoju repozytorium. Nawet wtedy należy jednak przypominać społeczności pracowników i studentów o jego istnieniu. Naturalną konsekwencją udostępniania treści w repozytorium jest ich wykorzystanie, na przykład w dydaktyce; służy ono nie tylko autorom, lecz także – w ramach danej instytucji – studentom czy doktorantom. Samym autorom umożliwia natomiast nie tylko wywiązanie się z nałożonych na nich zobowiązań, ale również szerokie udostępnienie prac, zapewnienie ich widoczności w Internecie oraz archiwizację.

W wypadku, gdy utworzeniu i prowadzeniu repozytorium nie towarzyszy wdrażanie polityki otwartego dostępu, rola komunikacji w rozwoju zasobów repozytorium staje się kluczowa. Bez starannie zaplanowanej kampanii informacyjnej repozytorium nie będzie wykorzystywane ani przez autorów, ani przez czytelników. Działania promocyjne są zazwyczaj podejmowane przez pracowników odpowiedzialnych za prowadzenie repozytorium, warto jednak połączyć ich siły z pracą osób odpowiedzialnych za komunikację wewnętrzną i zewnętrzną instytucji. Pozwoli to zwiększyć skalę i zasięg podejmowanych działań, a w efekcie również ich skuteczność.

## **6.2. Polityka otwartości w działaniu**

Kiedy w repozytorium znajdzie się wystarczająco dużo treści, aby można było wykorzystywać je w komunikacji zewnętrznej, można rozpocząć nowy etap działań promocyjnych i PR-owych. Należy oczywiście kontynuować komunikację wewnętrzną, przy czym obok działań adresowanych do pracowników warto zwrócić szczególną uwagę na informacje dla studentów, dla których repozytorium może stać się wartościowym i wygodnym źródłem materiałów naukowych.

Na tym etapie kluczową kwestią jest włączenie tematyki związanej z otwartym dostępem w działania komunikacyjne i promocyjne, skierowane do mediów, potencjalnych studentów, naukowców z innych instytucji

---

<sup>116</sup> W komunikacji z pracownikami warto podkreślać, że przyjęcie przez instytucję polityki otwartości ułatwi im negocjacje z dużymi wydawcami, z których wielu zgadza się na umieszczanie artykułów w repozytorium instytucjonalnym, o ile autora zobowiązuje do tego instytucja.

czy przedsiębiorców. Bogate zasoby umożliwiają z jednej strony promocję samego repozytorium jako miejsca dostępu do bieżącego dorobku naukowego instytucji, z drugiej – uzupełnienie o otwarte treści dotychczas realizowanej komunikacji ukierunkowanej na pozycjonowanie ekspertów, informowanie o projektach lub zespołach badawczych, które odniosły sukces, otrzymanych przez naukowców nagrodach czy innych ważnych osiągnięciach. Repozytorium może być także ważnym źródłem inspiracji dla osób odpowiedzialnych za PR uczelni – w udostępnianych pracach mogą znaleźć m.in. pomysły na artykuły lub wywiady do mediów wewnętrznych oraz na tematy warte szerokiego komunikowania. Otwarte publikacje naukowe są również doskonałym materiałem do komunikacji w mediach społecznościowych, które wymagają częstego i regularnego zasilania treściami; linki do ważnych i ciekawych artykułów opatrzone hasłowym komentarzem mogą uzupełniać komunikację instytucji na Facebooku czy Twitterze.

Przykładowe tematy komunikacji związanej z otwartymi pracami dostępnymi w repozytorium są następujące:

- kolejne kamienie milowe w rozwoju repozytorium, np.: „mamy już 1000 (2000, 3000) prac w repozytorium”, „w ciągu pierwszego roku działania repozytorium trafiło do niego 100 doktoratów”;
- najbardziej popularne treści, np.: „najbardziej popularna praca to artykuł *Propozycje nowych zasad kategoryzacji międzynarodowych czasopism humanistycznych*, który przeglądano 3691 razy”;
- statystyki repozytorium, np.: „prace z naszego repozytorium pobierane są także przez użytkowników spoza Polski, wśród których najliczniejszą grupę stanowią Chińczycy”;
- pozycja repozytorium w międzynarodowych rankingach, np.: „nasze repozytorium zajmuje obecnie 538. pozycję w światowym rankingu repozytoriów Webometrics; w porównaniu z rokiem ubiegłym awansowało o 3 pozycje”;
- wydarzenia związane z rozwojem repozytorium, np.: „repozytorium naszej instytucji zostało zarejestrowane w OpenAIRE – europejskiej infrastrukturze otwartego dostępu”;
- ważne publikacje pracowników w otwartych czasopismach naukowych (bądź opublikowane w czasopismach zamkniętych, ale udostępnione w repozytorium instytucjonalnym), np. „w PLOS ONE, jednym z najbardziej prestiżowych otwartych czasopism naukowych na świecie, opublikowany został artykuł (...) oparty na badaniach zespołu (...)”;
- Open Access Week, czyli organizowane przez Scholarly Publishing and Academic Resources Coalition (SPARC) coroczne światowe święto otwartego dostępu, kiedy instytucje z całego świata organizują wydarzenia wspierające otwarty dostęp, takie jak akcje edukacyjne, seminaria, webinaria itp.<sup>117</sup>

Powyższa lista nie wyczerpuje oczywiście możliwości wykorzystywania otwartych treści, w szczególności tych udostępnianych w repozytorium, w działaniach komunikacyjnych instytucji, ale wskazuje jedynie ich możliwe kierunki.

Środowisko zwolenników otwartości w nauce prowadzi serwisy informacyjne, do których można nadsyłać informacje o nowych inicjatywach dotyczących otwartego dostępu do treści naukowych. Oprócz możliwości znalezienia się na ich łamach, serwisy te oferują także dostęp do materiałów przydatnych w działaniach promujących otwarty dostęp. Materiały te są zazwyczaj dostępne na licencjach CC, dzięki czemu można je

117 Na stronie Open Access Week <http://www.openaccessweek.org/> można zapoznać się z organizowanymi w trakcie OAW wydarzeniami i rejestrować planowane wydarzenia własne.



bez przeszkód wykorzystywać we własnych działaniach.

Otwartanauka.pl informuje o najważniejszych wydarzeniach związanych z otwartym dostępem do treści naukowych w Polsce i na świecie, udostępnia wywiady z ekspertami z Polski i z zagranicy, rekomenduje narzędzia ułatwiające wdrażanie otwartych modeli. Serwis prowadzony jest przez ICM UW.

Uwolnijnaukę.pl to strona redagowana przez przedstawicieli instytucji należących do Koalicji Otwartej Edukacji. Znajdują się na niej informacje o bieżących wydarzeniach, a także sekcje z materiałami informacyjnymi i linkami do przydatnych źródeł.

Serwis Platformy Otwartej Nauki (pon.edu.pl) oferuje dostęp do licznych materiałów w postaci książek, raportów, materiałów wideo, materiałów informacyjnych i promocyjnych itp., poświęconych otwartości w nauce. Znajdują się w nim również informacje o planowanych wydarzeniach – konferencjach, szkoleniach i seminariach organizowanych dla społeczności akademickiej.

### **6.3. Podsumowanie**

Działania komunikacyjne są integralną częścią wdrażania otwartego dostępu w instytucji naukowej. Ważna jest zarówno komunikacja wewnętrzna, ułatwiająca sprawne wprowadzenie nowych rozwiązań i ich przyjęcie przez pracowników, jak i komunikacja zewnętrzna, umożliwiająca realizację części korzyści płynących z otwartości. Działania te powinny być prowadzone przy ścisłej współpracy władz instytucji, osób odpowiedzialnych za jej komunikację i PR oraz osób zajmujących się funkcjonowaniem i rozwojem infrastruktury dystrybucji treści, w szczególności repozytoriów.

## Bibliografia

A. Acharya, *Indexing Repositories: Pitfalls and Best Practices*, <http://www.or2015.net/wp-content/uploads/2015/06/or-2015-anurag-google-scholar.pdf>.

A. Apps, *Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata*, Dublin Core Metadata Initiative, 2005, <http://dublincore.org/documents/dc-citation-guidelines/>.

bepress, *Search Engine Optimization: How to Attract More Visitors to Your Repository*, Digital Commons Reference Material and User Guides. Paper 35. 2012, <http://digitalcommons.bepress.com/reference/35>.

B.-Ch. Björk, M. Laakso, P. Welling, P. Paetau, *Anatomy of green open access*, „Journal of the Association for Information Science and Technology” 2014, 65.2, s. 237–250, [http://www.researchgate.net/profile/Mikael\\_Laakso/publication/259675765\\_Anatomy\\_of\\_green\\_open\\_access/links/02e7e52d4e398d7a8c000000.pdf](http://www.researchgate.net/profile/Mikael_Laakso/publication/259675765_Anatomy_of_green_open_access/links/02e7e52d4e398d7a8c000000.pdf).

S. Brin, L. Page, *Reprint of: The anatomy of a large-scale hypertextual web search engine*, 2012, „Computer networks” 56.18, s. 3825–3833, <http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf>.

J. Brutlag, *Speed Matters*, Google Research Blog, 23 lipca 2009, <http://googleresearch.blogspot.com/2009/06/speed-matters.html>.

Chitika Insights, *The Value of Google Result Positioning*, 2013, <http://cdn2.hubspot.net/hub/239330/file-61331237-pdf/ChitikaInsights-ValueofGoogleResultsPositioning.pdf>.

E. Collins, C. Milloy, G. Stone, *Guide to Open Access Monograph Publishing for Arts, Humanities and Social Science Researchers*, OAPEN-UK, 2015, <http://oapen-uk.jiscbooks.org/files/2015/07/Guide-to-open-access-monograph-publishing-for-researchers-final.pdf>.

M. Fedoryszak, D. Tkaczyk, Ł. Bolikowski, *Large scale citation matching using Apache Hadoop*, „Research and Advanced Technology for Digital Libraries”, Springer Berlin Heidelberg, 2013, s. 362–365, <http://arxiv.org/pdf/1303.6906.pdf>.

Google, *Turndown Information for PageSpeed Service*, blog Google Developers, 3 sierpnia 2015, <https://developers.google.com/speed/pagespeed/service/Deprecation>.

S. Greenstein, F. Zhu, *Do Experts or Collective Intelligence Write with More Bias? Evidence from Encyclopædia Britannica and Wikipedia*, Harvard Business School Working Paper 2014, 15–023, [http://www.hbs.edu/faculty/Publication%20Files/15-023\\_145191c4-220f-4dd8-8d41-ee2d1f693716.pdf](http://www.hbs.edu/faculty/Publication%20Files/15-023_145191c4-220f-4dd8-8d41-ee2d1f693716.pdf).

S. Harnad, T. Brody, F. Vallieres, L. Carr, S. Hitchcock, Y. Gingras, Ch. Oppenheim, H. Stamerjohanns, E. R. Hilf, *The access/impact problem and the green and gold roads to open access*, „Serials review” 2004, 30.4, s. 310–314, <http://www.ost.qc.ca/Portals/0/docs/articles/2004/science.pdf>.

A. Holl, prezentacja konferencyjna *OA mandate of the Hungarian Academy of Sciences – Features and Experiences so Far*, <http://www.konkoly.hu/staff/holl/Athens.pdf>.

*Jak poprawić widoczność rekordów w Google Scholar*, <https://otwartanauka.pl/analysis/rozwiazania/jak-poprawic-widoczosc-rekordow-w-google-scholar>.

M. Kemman, M. Kleppe, S. Scagliola, *Just Google It* [w:] C. Mills, M. Pidd, E. Ward, *Proceedings of the Digital Humanities Congress 2012*. „Studies in the Digital Humanities”, Sheffield: HRI Online Publications, 2014, <http://www.hrionline.ac.uk/openbook/chapter/dhc2012-kemman>.

N.Y. Kiat, *Citation parsing using maximum entropy and repairs.*, Tech. rep., National University of Singapore, 2005, <http://bibliopediaut.googlecode.com/svn-history/r73/Bibliopedia/ParsCit/doc/yongKiatNgThesis.pdf>.

*Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce*, [http://www.nauka.gov.pl/g2/oryginal/2015\\_10/9f62cc350837b942e51ae23dd1f23df8.pdf](http://www.nauka.gov.pl/g2/oryginal/2015_10/9f62cc350837b942e51ae23dd1f23df8.pdf).

F. Lardinois, *Google Shuts Down PageSpeed Service for Accelerating Websites*, blog TechCrunch, 6 maja 2015, <http://techcrunch.com/2015/05/06/google-shuts-down-pagespeed-service-for-accelerating-websites/>.

E.D. Lopez-Cozar, N. Robinson-García, D. Torres-Salinas, *Manipulating Google Scholar citations and Google Scholar metrics: Simple, easy and tempting*, arXiv preprint arXiv:1212.0638 (2012). Preprint dostępny pod adresem: <http://arxiv.org/ftp/arxiv/papers/1212/1212.0638.pdf>.

P. Manghi, N. Manola, W. Horstmann, D. Peters, *An infrastructure for managing EC funded research output-The OpenAIRE Project*, „The Grey Journal (TGJ): An International Journal on Grey Literature”, 2010, 6.1, <https://pub.uni-bielefeld.de/publication/1972842>.

A. Mas-Bleda, M. Thelwall, K. Kousha, I.F. Aguillo, *Do highly cited researchers successfully use the social web?*, „Scientometrics” 2014, 101, <http://www.scit.wlv.ac.uk/~cm1993/papers/DoHCRSSuccessfullySocialWeb.pdf>.

B. Matthies, *Digital Commons Annual Report 2011*. Butler University Library, 2011, <http://digitalcommons.butler.edu/cgi/viewcontent.cgi?article=1003&context=reports>.

D. Mituzas, *Re: How did Wikipedia manage to get such a high Google PageRank?*, Quora, ostatnio modyfikowany 2012-07-25, <https://www.quora.com/How-did-Wikipedia-manage-to-get-such-a-high-Google-PageRank>.

*Monographs and Open Access. A report to HEFCE*, London 2015, <http://www.hefce.ac.uk/pubs/rereports/year/2015/monographs/>.

E. Orduña-Malea, J.M. Ayllón, A. Martín-Martín, E.D. López-Cózar, *About the size of Google Scholar: playing the numbers*. arXiv preprint arXiv:1407.6239, 2014, <http://arxiv.org/abs/1407.6239>.

*Otwarta Nauka w Polsce 2014. Diagnoza*, red. J. Szprot, Warszawa 2014, Wydawnictwa ICM, [http://repozytorium.ceon.pl/bitstream/handle/123456789/5330/Otwarta%20nauka%20w%20Polsce%202014\\_Diagnoza.pdf](http://repozytorium.ceon.pl/bitstream/handle/123456789/5330/Otwarta%20nauka%20w%20Polsce%202014_Diagnoza.pdf).

E. Priest, *Copyright and the Harvard Open Access Mandate*, „Northwestern Journal of Technology and Intellectual Property” 2012, vol. 10.

E. Rodrigues, *Open Access to Publications and Research Data in Horizon 2020: What Are the Requirements and How Can Institutional Repositories and OpenAIRE Help to Meet Them*, „Digital Presentation and Preservation of Cultural and Scientific Heritage”, 2014, IV, s. 353–360, <http://sci-gems.math.bas.bg:8080/jspui/bitstream/10525/2404/1/353-360-trDiPP2014-45-E.Rodrigues-final.pdf>.

D.J. Solomon, B.-C. Björk, (2012), *A study of open access journals using article processing charges*, „J. Am. Soc. Inf. Sci.”, nr 63: s. 1485–1495. doi: 10.1002/asi.22673.

D. Tkaczyk, Ł. Bolikowski, A. Czczeko, K. Rusek, *A modular metadata extraction system for born-digital articles*, „Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on. IEEE”, 2012, <http://www.depot.ceon.pl/bitstream/handle/123456789/1972/das1.pdf>.

D. Tkaczyk, P. Szostek, M. Fedoryszak, P.J. Dendek, Ł. Bolikowski, *CERMINE: automatic extraction of structured metadata from scientific literature*, „International Journal on Document Analysis and Recognition (IJ DAR)”, 2015: s. 1–19, <http://link.springer.com/article/10.1007/s10032-015-0249-8/fulltext.html>.

P. Thirion et. al., *ORBi in orbit, a user-oriented IR for multiple wins: why scholars take a real part in the success story*, <http://orbi.ulg.ac.be/handle/2268/65254>.

*The STM Report*, 4th edition, [http://www.stm-assoc.org/2015\\_02\\_20\\_STM\\_Report\\_2015.pdf](http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf).

K. Shearer, *Aligning Repository Networks and the Confederation of Open Access Repositories (COAR)*, 2015, <http://ir.inflibnet.ac.in/bitstream/1944/1886/1/47.pdf>.

K. Siewicz, *Otwarty dostęp do publikacji naukowych. Kwestie prawne*, WUW, Warszawa 2012, <http://pon.edu.pl/index.php/nasze-publicacje?pubid=12>.

S. Stanisławska-Kloc, *Polskie prawo autorskie dla bibliotek cyfrowych*, <http://www.bn.org.pl/download/document/1279799669.ppt>.

Stanowisko Prezydium KRASP i Prezydium PAN z dnia 5 lipca 2013 r. w sprawie zasad otwartego dostępu do treści publikacji naukowych i edukacyjnych, [http://www.aktualnosci.pan.pl/images/stories/pliki/2013/Wiadomości\\_biez/stanowisko\\_KRASP-PAN\\_open\\_access.pdf](http://www.aktualnosci.pan.pl/images/stories/pliki/2013/Wiadomości_biez/stanowisko_KRASP-PAN_open_access.pdf).

M. Starczewski, T. Lewandowski, *Po czym rozpoznać dobre repozytorium?*, „Biblioteka i Edukacja”, nr 6/2014, [http://www.depot.ceon.pl/bitstream/handle/123456789/6316/Po%20czym%20rozpozna%C4%87%20dobre%20repozytorium\\_Biblioteka%20i%20Edukacja.pdf](http://www.depot.ceon.pl/bitstream/handle/123456789/6316/Po%20czym%20rozpozna%C4%87%20dobre%20repozytorium_Biblioteka%20i%20Edukacja.pdf).

Ch. Stohn, *A mandate for Open Access: The University of Liège (ULg) and ULg Library*, <http://initiatives.exlibrisgroup.com/2013/05/a-mandate-for-open-access-university-of.html>.

P. Suber, *Otwarty dostęp*, przeł. R. Bogacewicz et. al., WUW, Warszawa 2014, <http://pon.edu.pl/index.php/nasze-publicacje?pubid=14>.

F. Summann i in., *COAR Roadmap Future Directions for Repository Interoperability*, „Procedia Computer Science” 2015, 52, s. 13–19, [https://www.coar-repositories.org/files/Roadmap\\_final\\_formatted\\_20150203.pdf](https://www.coar-repositories.org/files/Roadmap_final_formatted_20150203.pdf).

M. Teplitskiy, G. Lu, E. Duede, *Amplifying the Impact of Open Access: Wikipedia and the Diffusion of Science*, arXiv preprint arXiv:1506.07608 (2015). <http://arxiv.org/ftp/arxiv/papers/1506/1506.07608.pdf>.

Zarządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 20 marca 2015 r. w sprawie powołania Zespołu doradczego do spraw otwartego dostępu do treści naukowych, [http://www.nauka.gov.pl/g2/oryginal/2015\\_05/f0061d2ae21e462a5816f8a8cbe4fdfb.pdf](http://www.nauka.gov.pl/g2/oryginal/2015_05/f0061d2ae21e462a5816f8a8cbe4fdfb.pdf).

## Spis ilustracji

- Rysunek 1. Pierwsza działająca strona WWW info.cern.ch odtworzona dzięki dostępnemu on-line symulatorowi przeglądarki line-mode – s. 56
- Rysunek 2. Występowanie danych bibliograficznych w rekordzie: plik HTML i podlinkowany do niego plik PDF – s. 58
- Rysunek 3. Przykład błędnego użycia pola DC.creator – rekord na stronie repozytorium – s. 67
- Rysunek 4. Przykład błędnego użycia pola DC.creator – tagi <meta> w sekcji head pliku HTML – s. 68
- Rysunek 5. Przykład skutków błędnego użycia pola DC.creator – strona z wynikami w Google Scholar – s. 68
- Rysunek 6. Przykład skutków błędnego użycia pola DC.creator – błędne cytowanie przygotowane automatycznie przez Google Scholar – s. 69
- Rysunek 7. Pole dcTERMS.bibliographicCitation przystosowane do automatycznego odczytania (machine-readable) – s. 70

## Spis tabel

- Tabela 1. Zestawienie metadanych z formatów Highwire i Dublin Core w kontekście zaleceń Google Scholar – s. 73

