

Leszek Szafrński

Uniwersytet Jagielloński, Biblioteka Jagiellońska, Kraków

Repozytoria instytucjonalne na świecie: próba charakterystyki

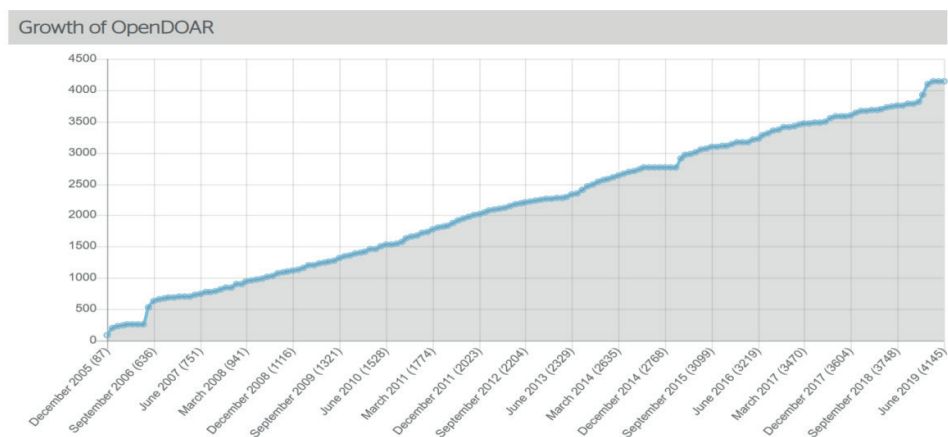
Streszczenie: Celem opisanego w artykule badania była analiza systemów obsługujących repozytoria instytucjonalne na uczelniach. Na potrzeby badania wykorzystano metodę analizy i krytyki piśmiennictwa oraz metodę wielokrotnego studium przypadku. Do badania zostało wybranych dwadzieścia repozytoriów z największą liczbą dokumentów zaindeksowanych w Google Scholar według portalu Ranking Web of Repositories. Badana była funkcjonalność repozytoriów, budowa i rodzaj systemu użytego do zarządzania repozytorium oraz kolekcje i typy udostępnianych zasobów. Przeprowadzone analizy pozwoliły na ustalenie podobieństw i różnic w funkcjonowaniu repozytoriów. W przyszłości mogą posłużyć do praktycznego doskonalenia funkcjonalności repozytoriów i rozwoju dalszych badań związanych z zarządzaniem i dystrybucją informacji naukowej.

Słowa kluczowe: funkcjonalność repozytoriów, Ranking Web of Repositories, repozytoria instytucjonalne, systemy repozytoryjne

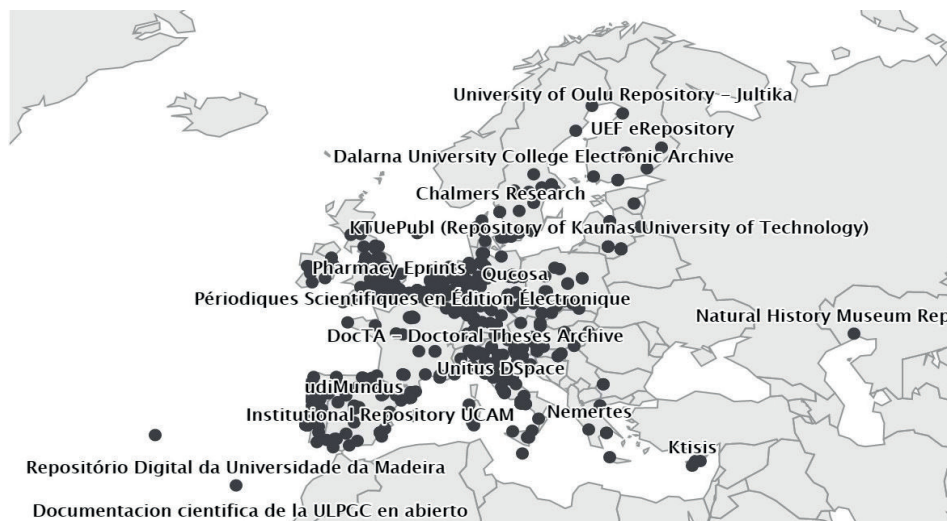
1. Wstęp

Repozytoria cyfrowe funkcjonują już od wielu lat. Pierwsze zostało założone przez Paula Ginsparga w 1991 r. (Ginsparg 2011, s. 4). W katalogu repozytoriów OpenDOAR notowane są od 2005 r. i obecnie jest w nim zgłoszonych 4145 repozytoriów (rys. 1), w tym 3576 instytucjonalnych (OpenDOAR Statistics 2019).

W Registry of Open Access Repositories (<http://roar.eprints.org>) zgłoszonych jest 4727 repozytoriów, w tym 3248 instytucjonalnych (Registry of Open Access Repositories 2019). Najwięcej repozytoriów znajduje się w Europie. Według OpenDOAR jest ich 1904 (OpenDOAR 2019, zakładka: Search, wyszukiwarka: Browse by Country). Zilustrowano to na rysunku 2.



Rys. 1. Liczba repozytoriów na świecie
(źródło: OpenDOAR Statistics 2019)



Rys. 2. Repozytoria w Europie
(źródło: OpenAIRE 2019)

Przedstawione dane liczbowe wskazują na to, że w kontekście zmian w komunikacji naukowej, jakie można zauważyć na świecie, repozytoria stały się nowym sposobem dostępu do naukowych zasobów cyfrowych. Widoczne jest to szczególnie w *Planie S*

wdrażanym przez cOAlition S, gdzie wymienia się repozytoria jako równoległy kanał publikowania wyników badań w otwartym dostępie i zachęca do archiwizowania wszystkich publikacji w repozytoriach, niezależnie od tego, gdzie pierwotnie zostaną wydane (cOAlition S 2019).

Jednak samo wdrożenie i uruchomienie repozytorium na uczelni jest dużym wyzwaniem (Szafrąński 2019, s. 133, 137). Dotyczy to zarówno technicznego aspektu wdrożenia systemu, jak i prac koncepcyjnych związanych z wyborem systemu informatycznego oraz projektowaniem: wyglądu interfejsów, kolekcji i typów zasobów.

2. Stan badań

Systemy repozytoryjne były już przedmiotem różnych badań. Należy m.in. wymienić publikacje pod redakcją Burtona B. Callicotta, Davida Scherera i Andrew Wesolka (Callicot et al. 2016), autora niniejszego artykułu (Szafrąński 2019) oraz Simony da Rocha Weitzel (Weitzel 2019). W pierwszej kompleksowo zostały omówione poszczególne komponenty działania repozytoriów – od wyboru systemu, przez politykę udostępniania, po przykłady wdrożeń. W drugiej autor przeprowadził badania dotyczące zarządzania zasobami cyfrowymi w polskich repozytoriach. W trzeciej została przeprowadzona analiza brazylijskich repozytoriów pod kątem ich funkcjonalności i jakości. O funkcjach polskich repozytoriów w początkowym okresie ich powstawania pisał Remigiusz Sapa (Sapa 2009).

W literaturze światowej jest wiele publikacji dotyczących działania repozytoriów, jednak poza wymienionymi wyżej analizami polskich i brazylijskich repozytoriów nie ma zbyt wielu badań większej liczby repozytoriów jednocześnie, szczególnie tych z pierwszych miejsc w rankingach światowych, i brakuje analiz z obszaru zastosowań praktycznych.

3. Cel i metoda badań

Celem badań była analiza systemów, które służą do zarządzania repozytoriami instytucjonalnymi działającymi na uczelniach, oraz analiza zasobów cyfrowych w nich gromadzonych. W konsekwencji posłużyła ona do rozpoznania unikalnych rozwiązań zastosowanych w poszczególnych repozytoriach i może w przyszłości pomóc w ulepszeniu funkcjonowania istniejących repozytoriów lub uruchamianych nowych systemów repozytoryjnych.

Do analizy zostało wybranych dwadzieścia instytucjonalnych repozytoriów akademickich rozumianych jako systemy do zarządzania zasobami cyfrowymi pracowników

danej uczelni (problemy terminologiczne zostały omówione w wielu publikacjach, np. Szafrński 2019) (tab. 1). Repozytoria zostały zakwalifikowane do badania na podstawie Ranking Web of Repositories. Autor wybrał systemy, w których zgromadzone zasoby mają największy wskaźnik indeksowania w Google Scholar (tab. 2) i które działają na uczelniach.

Tabela 1
Repozytoria wybrane do badania – zestawione alfabetycznie

Nazwa repozytorium	Państwo	Adres URL
Elektronna biblioteka BГУ [Электронная библиотека БГУ]	Białoruś	http://elib.bsu.by
EScholarship University of California	Stany Zjednoczone	https://escholarship.org
FGV Digital Repository	Brazylia	http://bibliotecadigital.fgv.br/dspace
HELDA University of Helsinki	Finlandia	https://helda.helsinki.fi
Institutional Repository of NCTU	Tajwan, Chiny	https://ir.nctu.edu.tw
Iowa State University Digital Repository	Stany Zjednoczone	https://lib.dr.iastate.edu
Kyoto University Research Information Repository	Japonia	https://repository.kulib.kyoto-u.ac.jp
LUME Repositório Digital UFRGS	Brazylia	https://lume.ufrgs.br
Lund University Publications	Szwecja	https://lup.lub.lu.se/search
MIT Libraries / DSpace@MIT	Stany Zjednoczone	https://dspace.mit.edu
Osaka University Knowledge Archive	Japonia	https://ir.library.osaka-u.ac.jp/repo/ouka/all
Repositório da Produção Científica e Intelectual da Unicamp	Brazylia	http://repositorio.unicamp.br
Repositório Institucional da UFSC	Brazylia	https://repositorio.ufsc.br
Repositório Institucional UNESP	Brazylia	https://repositorio.unesp.br
Repository Universitas Gadjah Mada	Indonezja	https://repository.ugm.ac.id
ScholarsArchive @ OSU	Stany Zjednoczone	https://ir.library.oregonstate.edu
The Digital Library of Theses and Dissertations of the University of São Paulo	Brazylia	http://www.teses.usp.br
Universitas Airlangga Repository	Indonezja	http://repository.unair.ac.id
UPCommons	Hiszpania	https://upcommons.upc.edu
Utrecht University Repository	Holandia	https://dspace.library.uu.nl

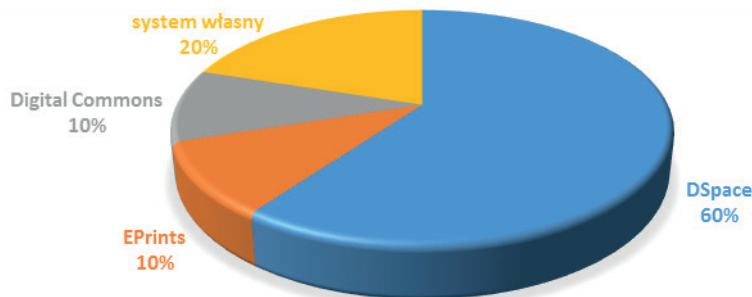
(źródło: Ranking Web of Repositories 2019)

Badania zostały przeprowadzone metodą wielokrotnego studium przypadku. W trakcie badań autor dokonał analizy systemów repozytoryjnych pod kątem używanego systemu informatycznego, liczby udostępnionych zasobów, rodzajów dostępnych kolekcji, na które podzielone są zasoby dostępne w systemie, typów zasobów, mechanizmów wyszukiwania informacji, integracji ORCID z repozytorium oraz polityki udostępniania.

Metoda analizy i krytyki piśmiennictwa posłużyła do przeglądu piśmiennictwa związanego z badaną tematyką.

4. Repozytoria instytucjonalne – analiza wyników badań

Większość badanych repozytoriów do zarządzania zasobami cyfrowymi używa oprogramowania DSpace (dwanaście repozytoriów). W czterech uczelniach używane są systemy własne. Pozostałe wykorzystywane systemy to EPrints (dwa repozytoria) i Digital Commons (dwa repozytoria). Dane przedstawiono na rysunku 3.



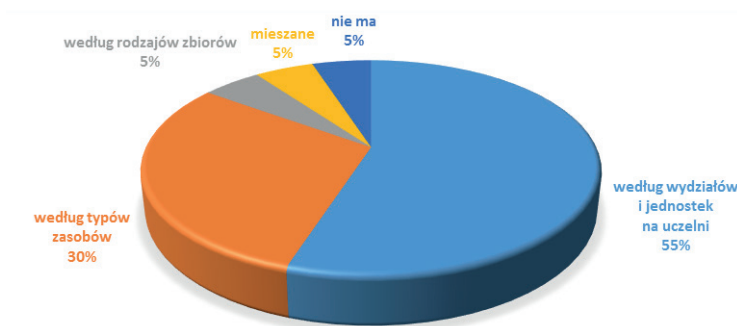
Rys. 3. Systemy do zarządzania zasobami w badanych repozytoriach

Dwa badane repozytoria są częścią większych systemów typu CRIS (*Current Research Information System*). Posiadają one dodatkowe funkcje, np. profile naukowców czy dane parametryczne (Joint 2008, s. 571–572).

W większości badanych repozytoriów zasoby podzielone są na główne kolekcje według struktury wydziałów i jednostek funkcjonujących na danej uczelni (jedenaście repozytoriów). Podobnie kolekcje zorganizowane są w większości polskich repozytoriów (Szafrąński 2019, s. 123).

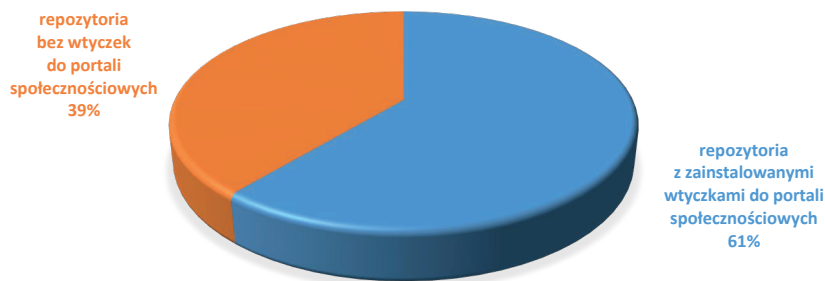
Innymi sposobami podziału głównych kolekcji są podziały według: typów zasobów (sześć repozytoriów), rodzajów zbiorów – nauka, dydaktyka, dziedzictwo

(jedno repozytorium), wydziałów i jednostek oraz typów zasobów (jedno repozytorium). W jednym repozytorium nie ma wyszczególnionych żadnych kolekcji. Dane procentowe przedstawione są na rysunku 4.



Rys. 4. Kolekcje główne w badanych repozytoriach

W jedenastu repozytoriach zainstalowane są wtyczki do portali społecznościowych. W pozostałych repozytoriach nie ma takiej funkcjonalności. Dane przedstawiono na rysunku 5.



Rys. 5. Używanie wtyczek do portali społecznościowych w badanych repozytoriach

Porównanie repozytoriów pod względem liczby publikacji indeksowanych w Google Scholar i liczby publikacji faktycznie znajdujących się w systemach repozytoryjnych wskazuje na duże rozbieżności. W wyniku analizy tabeli 2 można stwierdzić, że Google Scholar nie indeksuje wszystkich publikacji znajdujących się w repozytoriach.

W niektórych przypadkach rozbieżności są bardzo duże. Różnice dochodzą nawet do ponad 100 tys. publikacji w przypadku EScholarship University of California. Trudno wymienić wszystkie przyczyny. Najważniejsza, która wymieniana jest na stronach Google Scholar dla webmasterów, to posiadanie najnowszej wersji systemu repozytoryjnego. Do zarządzania repozytorium firma Google rekomenduje systemy: EPrints, Digital Commons, DSpace (Google Scholar 2019). Większość repozytoriów wymienionych w tabeli 2 działa na starszych wersjach podanych wyżej programów.

Dokumenty deponowane w repozytoriach powinny być zapisane w formacie PDF, a wielkość pliku nie powinna przekraczać 5 MB. Powinny być przygotowane według schematu: tytuł na początku większą czcionką niż reszta tekstu, pod tytułem powinien znaleźć się autor lub autorzy, na końcu tekstu podpisana w nagłówku bibliografia (Google Scholar 2019). W wielu tekstach naukowych nie przestrzega się jednak tych zasad. Jest to kolejny powód słabego indeksowania zasobów w Google Scholar.

Są trzy repozytoria, z których Google Scholar indeksuje więcej dokumentów, niż znajduje się w systemie: FGV Digital Repository, Utrecht University Repository i The Digital Library of Theses and Dissertations of the University of São Paulo. W przypadku FGV Digital Repository można to łatwo wytłumaczyć, ponieważ jest ono częścią większego systemu, do którego podpięte są jeszcze portale obsługujące czasopisma oraz książki. W rankingu podana jest suma indeksowanych w Google Scholar dokumentów z tych trzech baz. Natomiast w pozostałych dwóch repozytoriach trudno stwierdzić przyczynę takiej rozbieżności. Można przypuszczać, że repozytoria mają ukryte zasoby, ale są to tylko domysły autora nieoparte badaniami.

Tabela 2

Repozytoria zestawione według liczby publikacji indeksowanych w Google Scholar

Nazwa repozytorium	Liczba publikacji indeksowanych w Google Scholar	Liczba publikacji w repozytorium
LUME Repositório Digital UFRGS	153 000	191 845
Kyoto University Research Information Repository	134 000	195 910
Elektronnaja biblioteka BGU [Электронная библиотека БГУ]	109 000	164 815
EScholarship University of California	103 000	226 532
Repository Universitas Gadjah Mada	85 600	94 681
Utrecht University Repository	78 700	62 089

Tabela 2 cd.

Nazwa repozytorium	Liczba publikacji indeksowanych w Google Scholar	Liczba publikacji w repozytorium
The Digital Library of Theses and Dissertations of the University of São Paulo	77 600	52 603
Lund University Publications	70 200	189 031
Repositório Institucional UNESP	66 700	140 693
Institutional Repository of NCTU	66 500	148 316
FGV Digital Repository	65 600	23 256
Universitas Airlangga Repository	65 300	76 854
UPCommons	64 200	125 274
MIT Libraries / DSpace@MIT	61 600	106 263
Repositório da Produção Científica e Intelectual da Unicamp	61 600	141 188
Iowa State University Digital Repository	61 300	83 362
ScholarsArchive @ OSU	59 500	70 195
Repositório Institucional da UFSC	54 900	117970
Osaka University Knowledge Archive	54 500	66 905
HELDA University of Helsinki	53 800	80 456

(źródło: opracowanie na podstawie Ranking Web of Repositories i badań własnych)

W większości wymienionych w tabeli 2 repozytoriów część zasobów nie posiada pełnych tekstów dokumentów. Jest to w głównej mierze zależne od polityki otwartości przyjętej na danej uczelni. Wśród badanych repozytoriów tylko sześć przyjęło tzw. otwarty mandat (otwarty mandat – obowiązek deponowania przez autorów własnych publikacji w otwartym dostępie, często narzucony przez pracodawcę, instytucję naukową lub przyjęty jednogłośnie przez daną grupę autorów (Siewicz 2012, s. 13)). Pozostałe repozytoria działają na zasadach dobrowolności. W przypadku Biblioteki Cyfrowej Białoruskiego Uniwersytetu Państwowego (Электронная библиотека БГУ) nie udało się autorowi dotrzeć do dokumentów związanych z tym tematem.

W badanych repozytoriach nie ma rozbudowanych funkcji wspierających naukowca w cytowaniu prac. Tylko jedno repozytorium pozwala na pobranie metadanych do

cytowania i można to zrobić w ośmiu stylach cytowań, jednak sama funkcjonalność dostępna jest w innym systemie połączonym linkami z systemem repozytorium. W pięciu repozytoriach jest odesłanie do systemów zewnętrznych pozwalających na zarządzanie własną bibliografią, np. Mendeley czy EndNote.

Wyświetlanie ID i powiązanie nazwiska naukowca z bazą ORCID wspierane jest tylko w jednym repozytorium (rys. 6). Dodatkowo to repozytorium powiązane jest z innymi systemami uczelni, np. bazą profili naukowców.



Rys. 6. Wspieranie ORCID w repozytoriach

W badanych repozytoriach obok standardowych typów zasobów takich jak książki, artykuły i rozdziały występują również unikatowe, rzadko spotykane w repozytoriach typy publikacji, np. raport techniczny, eksperyment, działowy biuletyn informacyjny, wykład, adnotacja. Przepuszczalnie wymyślono je na potrzeby konkretnych, specyficznych dla danego repozytorium zasobów i w związku z tym nie do końca przystają one do obowiązujących standardów.

W poszczególnych repozytoriach są znaczne różnice w liczbie typów zasobów. Przedstawia to tabela 3. Rozbieżność jest bardzo duża – od podziału na trzy typy zasobów do nawet siedemdziesięciu ośmiu, co wskazuje na dostosowywanie typologii zasobów do konkretnych potrzeb danej uczelni i funkcji, jaką ma spełniać repozytorium. Przykładowo zadaniem głównym biblioteki cyfrowej The Digital Library of Theses and Dissertations of the University of São Paulo jest archiwizowanie rozpraw magisterskich, doktorskich i habilitacyjnych.

Tabela 3
Repozytoria zestawione według liczby typów zasobów

Nazwa repozytorium	Liczba typów zasobów
Elektronnaja biblioteka BGU [Электронная библиотека БГУ]	78
Repositório da Produção Científica e Intelectual da Unicamp	73
LUME Repositório Digital UFRGS	36
UPCommons	35
ScholarsArchive @ OSU	27
HELDA University of Helsinki	24
Repositório Institucional UNESP	23
FGV Digital Repository	21
Universitas Airlangga Repository	19
Osaka University Knowledge Archive	16
Kyoto University Research Information Repository	14
Utrecht University Repository	14
EScholarship University of California	10
Repository Universitas Gadjah Mada	10
Lund University Publications	10
Institutional Repository of NCTU	7
The Digital Library of Theses and Dissertations of the University of São Paulo	3
Repositório Institucional da UFSC	3
MIT Libraries / DSpace@MIT	trudno określić
Iowa State University Digital Repository	trudno określić

Podobne są natomiast w repozytoriach mechanizmy wyszukiwania publikacji. W każdym systemie jest możliwość wyszukiwania prostego i zaawansowanego oraz tzw. discover, czyli zawężania wyników wyszukiwania według przyjętych filtrów. W jednym repozytorium ciekawostką jest, że przeszukiwać w systemie można tylko przez zewnętrzną wyszukiwarkę będącą częścią agregatora NARCIS. W dziewięciu repozytoriach można filtrować wyniki m.in. według czterech podstawowych kryteriów: autora, tytułu, typu zasobu oraz tematu. W dziewiętnastu repozytoriach można filtrować m.in. według autora. W jednym repozytorium są dwadzieścia trzy filtry do zawężania wyników. Ponadto każde z repozytoriów ma własne konfiguracje filtrów. Dostępne są w różnych ustawieniach, czyli każde repozytorium definiuje sobie własny

zestaw filtrów. Oprócz wymienionego wyżej zawężania wyników według kryteriów: autor, tytuł, typ zasobu i temat występują w różnych konfiguracjach kryteria: data, wydawca, język, wydziały, jednostki, specjalizacja, prawa dostępu, promotor, abstrakt, projekt, słowa kluczowe, abstrakt, obszar wiedzy, format pliku, obszar wiedzy itp.

5. Podsumowanie

Z całą pewnością można stwierdzić, że największym problemem repozytoriów jest indeksowanie w Google Scholar. Pomimo używania zalecanych przez Google systemów nie wszystkie publikacje są rozpoznawane poprawnie przez wyszukiwarkę. Oczywiście nie jest to jedyna trudność, ale jedna z podstawowych. Przyglądając się samym systemom, można zauważyć, że większość repozytoriów bazuje na systemach *open source* bez rozbudowywania ich funkcjonalności ponad to, co oferuje sam system. Różnice występują na pewno w systemach autorskich, które dostosowywane są już do konkretnych potrzeb. Na podstawie analizy liczby typów zasobów i nadawanych im nazw można stwierdzić, że w niektórych repozytoriach jest to przygotowywane dość chaotycznie. Inaczej wygląda to w przypadku głównych kolekcji, które dość wiernie odzwierciedlają strukturę uczelni.

Ciekawą funkcjonalnością, którą coraz częściej można zauważyć w repozytoriach, jest instalacja wtyczek do portali społecznościowych, które służą użytkownikom do szybkiego przekazania informacji o dostępnych zasobach. Dość zaskakujące jest to, że tak popularny na świecie identyfikator jak ORCID jest wspierany tylko w jednym repozytorium. Wyszukiwanie zasobów przez wyszukiwarki wewnętrzne działa podobnie we wszystkich repozytoriach.

Przyglądając się przeprowadzonym badaniom i bazując na własnym doświadczeniu, autor stwierdził, że badane repozytoria w większości przypadków są na wczesnym etapie rozwoju pomimo dużej ilości zasobów w nich zgromadzonych. Na liczbę zasobów ma też wpływ przyjęta polityka otwartości. Repozytoria, które przyjęły politykę otwartego mandatu, są na najwyższych miejscach w Ranking Web of Repositories.

Bibliografia

- Callicott B.B., Scherer D., Wesolek A. (2016). Making Institutional Repositories Work [online]. West Lafayette: Purdue University Press. e-ISBN 9781612494227. Dostępny w: <https://oopen.org/download?type=document&docid=626981> [przełączany 1.06.2019].
- cOAlition S (2019). Plan S: Making full and immediate Open Access a reality: principles and implementation [online]. Science Europe AISBL. Dostępny w: <https://www.coalition-s.org/principles-and-implementation> [przełączany 27.06.2019].

- Ginsparg P. (2011). It was twenty years ago today... [online]. Dostępny w: <https://arxiv.org/pdf/1108.2700.pdf> [przełączany 1.05.2019].
- Google Scholar (2019). Inclusion Guidelines for Webmasters [online]. Google Scholar. Dostępny w: <https://scholar.google.com/scholar/inclusion.html> [przełączany 28.06.2019].
- Joint N. (2008). Current Research Information Systems, Open Access Repositories and Libraries: ANTAEUS. *Library Review* [online], vol. 57, no. 8, s. 570–575. Dostępny w: <https://doi.org/10.1108/00242530810899559> [przełączany 17.06.2019].
- OpenAIRE Share publications (2019) [online]. OpenAIRE. Dostępny w: <https://explore.openaire.eu/participate/deposit-publications> [przełączany 17.06.2019].
- OpenDOAR Statistics (2019) [online]. JISC. Dostępny w: http://v2.sherpa.ac.uk/view/repository_visualisations/1.html [przełączany 13.06.2019].
- OpenDOAR zakładka: Search, wyszukiwarka: Browse by Country (2019) [online]. OpenDOAR, JISC. Dostępny w: https://v2.sherpa.ac.uk/view/repository_by_country [przełączany 17.06.2019].
- Ranking Web of Repositories (2019) [online]. Cybermetrics Lab – CSIC. Dostępny w: <http://repositories.webometrics.info/en/institutional> [przełączany 1.06.2019].
- Registry of Open Access Repositories (2019) [online]. University of Southampton. Dostępny w: <http://roar.eprints.org> [przełączany 17.06.2019].
- Sapa R. (2009). Repozytoria i spisy bibliograficzne publikacji pracowników uczelni w polskich bibliotekach akademickich. W: Biblioteki i ich klienci: między płatnym a bezpłatnym komunikowaniem się w erze zasobów cyfrowych i sieci [Libraries and their clients: free or fee services supporting social communication in digital era] [online]. Red. M. Kocójowa. Seria III, ePublikacje Instytutu Informacji Naukowej i Bibliotekoznawstwa, 1896–3846, nr 6. Kraków: Instytut Informacji Naukowej i Bibliotekoznawstwa, Uniwersytet Jagielloński, s. 213–223. ISBN 9788392159339. Dostępny w: <https://isi.uj.edu.pl/wydawnictwa-iinib-uj/seria-3/06> [przełączany 17.05.2019].
- Siewicz K. (2012). Otwarty dostęp do publikacji naukowych: kwestie prawne. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego. ISBN 9788323509677.
- Szafrąński L. (2019). Zarządzanie zasobami dokumentów elektronicznych w instytucjonalnych repozytoriach akademickich. Warszawa: Stowarzyszenie Bibliotekarzy Polskich. ISBN 9788365741257.
- Weitzel S. da R. (2019). O mapeamento dos repositórios institucionais brasileiros: perfil e desafios. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação* [online], vol. 24, no. 54, s. 105–123. Dostępny w: <https://doi.org/10.5007/1518-2924.2019v24n54p105> [przełączany 17.06.2019].