

Anna Małgorzata Kamińska
Instytut Bibliotekoznawstwa i Informatyki Naukowej
Uniwersytet Śląski w Katowicach
anna.kaminska@us.edu.pl

OpenCitations – otwarty indeks cytowań publikacji naukowych

Streszczenie: Mimo stopniowego upowszechniania się dokumentowania badań naukowych na zasadach otwartego dostępu oraz innych koncepcji związanych z „Nauką 2.0”, prace naukowe publikowane w ten sposób stanowią wciąż jedynie „uzupełniającą” działalność naukowców. Dzieje się tak zwłaszcza w przypadkach, gdy ewaluacja działalności naukowej oparta jest na „twardych” wskaźnikach bibliometrycznych wyliczanych na podstawie „uznanych” źródeł danych, które gromadzą zwykle informacje jedynie o pracach wydawanych w ramach „prestżowych” wydawnictw. Wyliczanie podobnych wskaźników dla wydawców/platform wspierających publikowanie zgodnie z koncepcjami „Nauki 2.0” jest obecnie mocno utrudnione z powodu problemów z obliczeniem wartości skumulowanej będącej wypadkową wartości wskaźników prezentowanych na konkretnych platformach. Wynika to z faktu, że każdy wydawca lub platforma gromadzi najczęściej jedynie pewien podzbiór prac danego autora, a ponieważ podzbiory te mogą nie być rozłączne, to proste zsumowanie liczby cytowań wartości cząstkowych publikowanych przez każdy z podmiotów nie pozwoli na obliczenie poprawnej wartości skumulowanej. Artykuł przedstawia inicjatywę OpenCitations, która korzystając z uznanych standardów, proponuje ontologię indeksu cytowań, według której gromadzi i udostępnia dane w postaci struktur grafów cytowań. Otwarte standardy oraz wolny dostęp do gromadzonych danych sprawiają, że inicjatywa OpenCitations stanowić może interesujący kierunek dla integracji różnych źródeł danych o cytowaniach, co pozwoli na łatwe obliczanie skumulowanych/globalnych wskaźników, zaś możliwość pozyskania całego korpusu zgromadzonych danych otworzy perspektywy do realizacji bardziej zaawansowanych bibliometrycznych badań własnych.

Słowa kluczowe: OpenCitations, indeks cytowań, bibliometria, źródła danych, otwarty dostęp

Wprowadzenie

Przyjmuje się, że początki koncepcji budowy indeksów cytowań datuje się na rok 1955, kiedy to Eugene Garfield opublikował w „Science”¹ propozycję tworzenia struktur danych opisujących cytowania artykułów naukowych (ang. *Citation indexes for science*) m.in. jako narzędzia oceny czasopism. Pierwszy *Science Citation Index* (SCI) ukazał się w 1963 r. i zawierał dane o 102 tys. artykułów, które zostały opublikowane w roku 1961 w 613 wybranych czasopismach. Choć początkowo rozwój analiz bibliometrycznych stymulowany był głównie potrzebą obserwacji rozwoju dziedzin nauki oraz oceny czasopism, to od pewnego już czasu ilościowe wskaźniki bibliometryczne stanowią również jedną ze składowych ocen parametrycznych instytucji naukowych oraz samych naukowców zarówno na świecie, jak i w Polsce. Jednak racjonalność decyzji dotyczących stymulacji rozwoju ośrodków naukowych w oparciu o wskaźniki ilościowe zależy od kompletności i wiarygodności źródeł danych, na podstawie których wskaźniki te są wyliczane.

¹ GARFIELD, E. Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science* 1955, Vol. 122, No. 3159, s. 108–111. ISSN 0036-8075.

Obserwować obecnie można rozwój indeksów cytowań w ramach dwóch przeciwległych biegunów. Pierwszy to źródła kontrolowanej jakości utrzymywane przez instytucje komercyjne, współpracujące bezpośrednio z wydawcami czasopism naukowych. Przykładem takiego indeksu jest z całą pewnością SCI, powołany do istnienia przez wspomnianego Eugene'a Garfielda, a utrzymywany i rozwijany obecnie przez przedsiębiorstwo Clarivate Analytics. Indeksowane informacje pochodzą bezpośrednio od wydawców wybranych czasopism naukowych, stanowią więc wiarygodne źródło wskaźników obliczanych dla czasopism czy naukowców. Jednak dostęp do danych ograniczony płatną subskrypcją, a jednocześnie możliwość pobierania jedynie niewielkich podzbiorów danych (wynikające z przyjętego modelu biznesowego) powodują, że realizacja własnych analiz bibliometrycznych bazujących na tym źródle jest mocno ograniczona. Warto również zwrócić uwagę, że profil indeksowanych czasopism ukierunkowany głównie na nauki twarde sprawia, że dla przedstawicieli nauk humanistycznych indeks ten nie może stanowić źródła wiarygodnych wskaźników.

Drugie, przeciwne podejście, implementowane przez autorów usługi Google Scholar, bazuje na pozyskiwaniu deklaracyjnych informacji wprost, poprzez analizę zawartości stron WWW instytucji naukowych oraz wydawnictw publikacji elektronicznych, na których naukowcy lub wydawcy umieszczają bądź to pełne treści swoich publikacje, bądź jedynie informacje bibliograficzne je opisujące (włączając w nie bibliografię załącznikową). Powoduje to, że wartości wskaźników prezentowane na stronach portalu Google Scholar nie mogą być traktowane bezkrytycznie, co oczywiście nie zmienia ogólnie wysokiej przydatności tej usługi w zastosowaniach wyszukiwania publikacji naukowych konkretnych autorów lub zadanym dziedzin. Warto zauważyć, że choć dostęp do danych jest w tym przypadku darmowy, to również istnieje ograniczenie co do wolumenu jednocześnie pobieranych danych, zaś sam format ich udostępniania nie sprzyja budowaniu na ich podstawie własnych trawersowalnych indeksów cytowań, na których oprócz by można bardziej zaawansowane bibliometryczne badania własne². Implikuje to czasem konieczność wykorzystania innych metod³ do rejestrowania danych bibliograficznych z autopsji, na bazie których możliwa jest realizacja wspomnianych badań, jak miało to miejsce w przypadku badań opisanych w innym opracowaniu autorki⁴.

Omawiając aspekty funkcjonalne istniejących indeksów cytowań, nie sposób pominąć ich oceny pod kątem dynamicznie ostatnio rozwijających się najnowszych sposobów komunikowania dorobku naukowego, jakimi są: samopublikowanie, publikowanie na zasadach swobodnego dostępu do treści czy uwzględnianie innych relacji wiążących poszczególne dokumenty i ich autorów niż tylko „cytowanie” lub „autorstwo” (np. relacje wykorzystywane w altmetrii). Chcąc utrzymać dotychczasową pozycję rynkową, obecni dostawcy usług związanych z analizą indeksów cytowań z całą pewnością zrewidować będą musieli w nie-

² KAMIŃSKA, A.M. ProBIT – prospektywna metoda tworzenia trawersowalnych indeksów cytowań a współczesne problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych. *Zagadnienia Informacji Naukowej* 2017, nr 1 (w druku).

³ KAMIŃSKA, A.M. Tam, gdzie zaczyna się bibliometria, czyli jak pozyskać materiał analityczny z autopsji. *Biuletyn EBIB* [online]. 2017, No 173. [Dostęp 16.08.2017]. ISSN 1507-7187. Dostępny w: <http://open.ebib.pl/ojs/index.php/ebib/article/view/534>.

⁴ KAMIŃSKA, A.M. Od druków źródłowych po mapy nauki. Bibliograficzna baza danych GRUBA. W: KO-WALSKA, M., OSIŃSKA, V. (red.). *Wizualizacja informacji w humanistyce*. Toruń: Wydaw. Uniwersytetu Mikołaja Kopernika, 2017 (w druku).

odległej przyszłości zarówno swoje modele ontologiczne, zgodnie z którymi dane są gromadzone i udostępniane, jak i biznesowe modele funkcjonowania. Jest to spowodowane obserwowanym już obecnie, nawet w krajach premiujących naukowców za publikowanie w „uznanych zamkniętych” wydawnictwach, powszechnym wzrostem zainteresowania otwartą nauką, i dzieje się tak również w Polsce. Kiedy wzrosną proporcje udziału otwartego rynku wydawniczego w całości obszaru dokumentowania badań (a jednym z punktów zwrotnych stać się tutaj może większa liczba „otwartych prac” cytowanych przez „prace zamknięte” nad liczbą relacji odwrotnej), ogólnodostępne dane bibliometryczne stać się mogą bardziej cenne od reglamentowanych obecnie danych pochodzących z wydawnictw zamkniętych. Podobnie jak w przypadku oprogramowania „otwartoźródłowego” (ang. *open source*), którego twórcom trudno kiedyś było myśleć o działalności komercyjnej i do działania nie motywowała ich chęć zysków materialnych, a dzisiaj oprogramowanie to staje się coraz bardziej popularne i wiarygodne (gdyż każdy, mając dostęp do kodów źródłowych, może prześledzić ich funkcjonowanie), zaś działalność komercyjną buduje się wokół usług udostępniania bardziej zaawansowanych funkcjonalności czy usług serwisowych, tak i udostępnianie podstawowych danych bibliograficznych na zasadach wolnego dostępu mogłoby uwiarygodnić wartości obliczanych współczynników bibliometrycznych, a wzrost zainteresowania tymi analizami mógłby spowodować wzrost popytu na usługi specjalistyczne wokół nich budowane.

Inicjatywa wolnego dostępu do cytowań (I4OC)

Budowanie ogólnodostępnych usług bibliograficznych wymaga istnienia ogólnodostępnych bibliograficznych źródeł danych. Mimo że większość wydawców publikacji naukowych przekazuje informacje o bibliografiach załącznikowych publikowanych przez nich artykułów do CrossRef w ramach uczestniczenia w usłudze „CrossRef Cited-by”, to domyślnie dostarczane dane nie są upubliczniane. Udostępnianie takich danych z poziomu tej usługi możliwe jest jedynie na wyraźne życzenie wydawców. Dotychczas udostępniano jedynie znikomy procent danych bibliograficznych, jednak od niedawna sytuacja ta zaczęła się gwałtownie zmieniać. Inicjatywa wolnego dostępu do cytowań⁵ (ang. The Initiative for Open Citations – I4OC), do której powstania przyczynił się Dario Taraborelli z fundacji Wikimedia, polegająca na współpracy środowisk wydawców, naukowców i innych zainteresowanych stron, w ciągu krótkiego czasu przekonała większość znaczących wydawnictw naukowych do udostępniania danych bibliograficznych przesyłanych do CrossRef, powodując na dzień dzisiejszy wzrost odsetka udostępnianych danych do 45%.

Wśród wydawców udostępniających na dzień dzisiejszy metadane swoich publikacji naukowych wymienić należy:

- American Geophysical Union,
- Association for Computing Machinery,
- BMJ,
- Cambridge University Press,
- Co-Action Publishing,
- Cold Spring Harbor Laboratory Press,
- Copernicus GmbH,
- eLife,

⁵ *Initiative for Open Citations* [online]. [Dostęp 16.08.2017]. Dostępny w: <https://i4oc.org>.

- EMBO Press,
- Faculty of 1000, Ltd.,
- Frontiers Media SA,
- Geological Society of London,
- Hamad bin Khalifa University Press (HBKU Press),
- Hindawi,
- International Union of Crystallography,
- Leibniz Institute for Psychology Information,
- MIT Press,
- PeerJ,
- Pensoft Publishers,
- Portland Press,
- Public Library of Science,
- Royal Society of Chemistry,
- SAGE Publishing,
- Springer Nature,
- Taylor & Francis Group,
- The Rockefeller University Press,
- The Royal Society,
- Ubiquity Press, Ltd.,
- Wiley.

OpenCitations

Jednym ze skutków wzrostu zainteresowaniem koncepcją wolnej nauki było powstanie otwartego indeksu cytowań nazwanego przez jego twórców OpenCitations Corpus (OCC) i udostępnianego jako usługa internetowa spod adresu: opencitations.net⁶. Trzeba wyraźnie odróżnić podmioty I4OC oraz OCC, gdyż ten ostatni jest inicjatywą polegającą na stworzeniu infrastruktury dedykowanej udostępnianiu danych o cytowaniach artykułów naukowych zgodnie z koncepcją *Link Open Data*, w formacie RDF (JSON-LD).

OCC rozpoczęło swoją działalność formalnie w roku 2010, z wizją globalnego zasięgu i misją zmiany krajobrazu procesu publikowania i komunikacji naukowej, gdyż było ukierunkowane na darmowe dostarczanie informacji o cytowaniach w sposób umożliwiających trawersowanie ścieżek cytowań za pomocą identyfikatorów URI.

W roku 2015 OpenCitations opracowało nowe modele gromadzenia danych bibliograficznych, nowe oprogramowanie i nowy sposób zautomatyzowanej akwizycji danych. Obecnie dane bibliograficzne pobierane są z Open Access Subset of PubMed Central⁷ z zastosowaniem interfejsu Europe PubMed Central, a następnie są przetwarzane i składowane w ramach infrastruktury Wydziału Informatyki i Inżynierii Uniwersytetu w Bolonii.

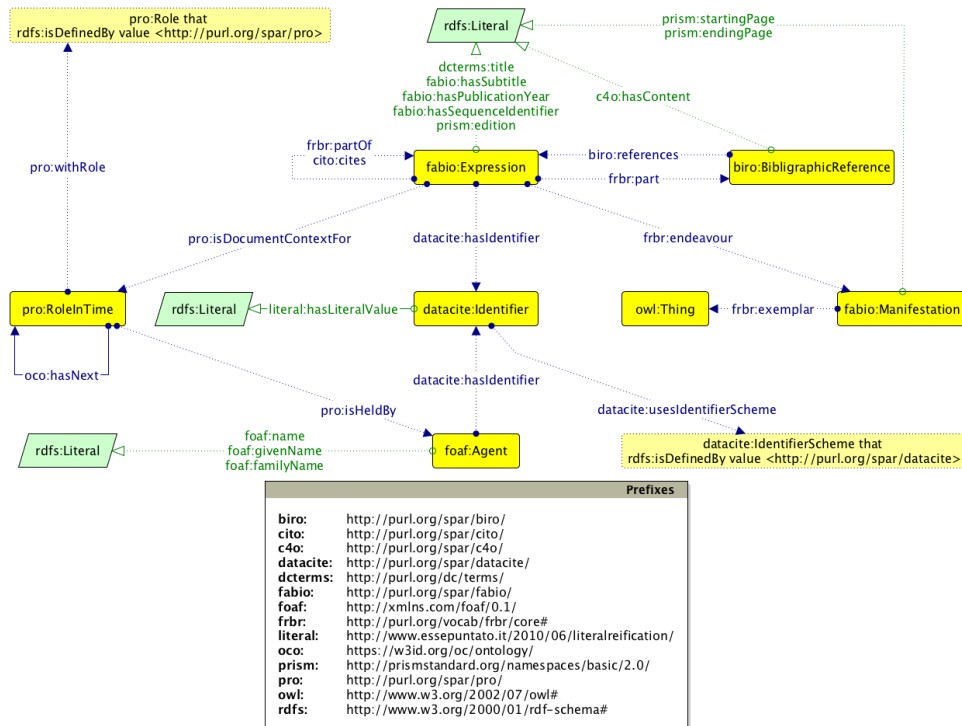
⁶ PERONI, S., SHOTTON, D., VITALI, F. One year of the OpenCitations Corpus: Releasing RDF-based scholarly citation data into the Public Domain. W: *Proceedings of the 16th International Semantic Web Conference (ISWC 2017)* [online]. [Dostęp 16.08.2017]. Dostępny w: <https://w3id.org/people/essepuntato/papers/oc-iswc2017.html>.

⁷ Europe PMC [online]. [Dostęp 16.08.2017]. Dostępny w: <http://europepmc.org/>.

15 września 2017 r. OCC zgromadził dane o 239 827 cytujących jednostkach bibliograficznych i zawiera informacje o 10 208 696 cytowaniach 5 590 615 jednostek cytowanych. Gdy tylko przetworzone zostaną informacje o wszystkich (ponad półtora miliona) artykułach z Open Access Subset of PubMed Central, rozpocznie się przetwarzanie informacji o cytowaniach z ponad 16 mln artykułów, które udostępnione zostały przez Crossref w odpowiedzi na działania I4OC oraz z artykułów, które zostaną do tego czasu udostępnione na skutek nieustających działań I4OC.

Model informacyjny

Nawet najbardziej kompletne dane będą stanowić jedynie bezużyteczny zbiór informacji, jeśli nie będą składowane w sposób umożliwiający ich odpowiednie wykorzystanie. W rytm rozwoju nauki i kształtowania się sposobów dokumentowania dorobku naukowego opracowywane były różne modele organizowania danych bibliograficznych. Twórcy platformy OpenCitations postanowili nie budować zamkniętych rozwiązań, a podążać za otwartymi standardami mogącymi służyć za referencyjne modele danych. W przypadkach potrzeby opracowywania modeli dla nowych obszarów zostały one odpowiednio udokumentowane i opublikowane⁸. Do gromadzenia danych o założonym zakresie informacyjnym konieczne było zintegrowanie kilku różnych ontologii opisujących poszczególne obszary dziedzinowe. Ilustruje to diagram przedstawiony na rys. 1.



Rys. 1. Diagram ontologii wykorzystywanej przez OpenCitations
Źródło: Model. W: *OpenCitations* [online]. [Dostęp 16.08.2017].
Dostępny w: <http://opencitations.net/model#refillustration0>.

⁸ PERONI, S., SHOTTON, D. Metadata for the OpenCitations Corpus [online]. 2017. [Dostęp 16.08.2017]. Dostępny w: <https://dx.doi.org/10.6084/m9.figshare.3443876>.

Rysunek stanowi schemat wykonany w narzędziu Grafoo⁹, które stworzone zostało specjalnie w celu umożliwienia graficznego modelowania języka OWL (ang. Web Ontology Language). Diagram ten, podobnie jak diagram klas języka UML omówiony np. w innej pracy autorki¹⁰, przedstawia perspektywę strukturalną, ukazując, jak powiązane są ze sobą poszczególne obszary dziedzinowe. Całość koncepcji opiera się na wykorzystaniu ontologii SPAR (ang. Semantic Publishing and Referencing) wraz z poszczególnymi jej elementami składowymi, z których zasadniczą funkcję pełnią:

- FaBiO (ang. FRBR-alignedBibliographic) – wykorzystywana do opisu wszystkich metadanych źródeł cytujących i cytowanych oraz form wydawniczych, w których się zawierają,
- PRO (ang. Publishing RolesOntology) – wykorzystywana do opisu ról podmiotów operujących na zasobach bibliograficznych (np. autorzy, wydawcy, redaktorzy),
- BiRO oraz C4O (ang. Bibliographic Reference Ontology/CitationCounting and ContextCharacterizationOntology) – wykorzystywane do opisu każdej z pozycji bibliografii załącznikowej jednostki cytującej,
- Datacite – do opisu wszystkich identyfikatorów (np. DOI, ISSN,...) jednostek bibliograficznych i operujących na nich podmiotach.

Tak skonstruowany model informacyjny, nazwany dla wygody OCO¹¹ (ang. OpenCitationsOntology), choć formalnie nie jest definicją nowej ontologii, to integruje w sobie informacje o jednostkach bibliograficznych (zarówno cytujących, jak i cytowanych), formach wydawniczych poszczególnych jednostek, rekordach bibliografii załącznikowych łączących jednostki cytowane z cytującymi, odpowiedzialnościach i identyfikatorach.

Funkcjonalność

Choć na pierwszy rzut oka platforma OpenCitations nie wyróżnia się ani szatą graficzną, ani bogactwem funkcjonalności, to warto zwrócić uwagę na zastosowany do jej budowy stos technologiczny. Wszystkie składowe oprogramowania dostępne są darmowo na zasadach licencji ISC lub GPLv2, a serce systemu stanowi zaawansowany i wysokoskalowalny silnik przetwarzania danych o strukturze sieciowej – Blazegraph¹². Wspiera on standard RDF¹³ (ang. Resource Description Framework) będący językiem opisu danych o strukturze sieciowej, ze składnią opartą na XML, a opracowanym przez konsorcjum W3C. Dane z systemu pozyskać możemy na kilka sposobów:

- poprzez ich przeglądanie za pomocą bardzo uproszczonego interfejsu z poziomu przeglądarki internetowej (np. <https://w3id.org/oc/corpus/br/1>),
- poprzez pobranie (<http://opencitations.net/download>) ze stron Figsharefragmentów korpusu (dane aktualizowane są w miesięcznych cyklach) jako archiwum ZIP, zawierające rekordy w formacie JSON dla poszczególnych obszarów danych (tj. dla osób, ich odpowiedzialności względem poszczególnych jednostek bibliograficz-

⁹ Grafoo: *Graphical Framework for OWL Ontologies* [online]. [Dostęp 16.08.2017]. Dostępny w: <http://www.es-sepunto.it/graffoo/>.

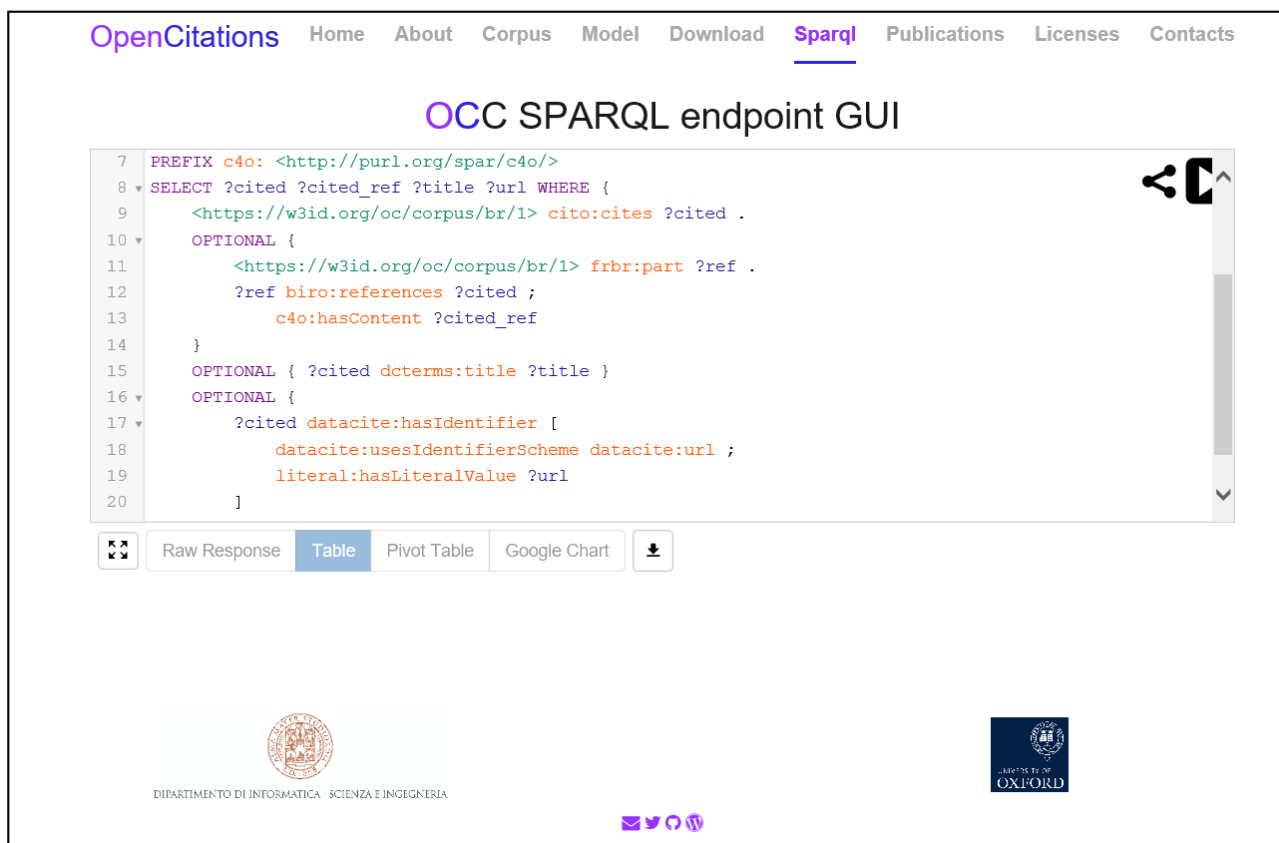
¹⁰ KAMIŃSKA, A.M. O rozwoju graficznych języków komunikacji. *Zagadnienia Informacji Naukowej* 2017, nr 2 (w druku).

¹¹ OCO, *the OpenCitations Ontology* [online]. OpenCitations, 2016. [Dostęp 16.08.2017]. Dostępny w: <http://opencitations.net/ontology.html>.

¹² Blazegraph [online]. [Dostęp 16.08.2017]. Dostępny w: <https://www.blazegraph.com/>.

¹³ RDF [online]. W3C, 2017. [Dostęp 16.08.2017]. Dostępny w: <https://www.w3.org/RDF/>.

- nych, jednostek bibliograficznych, rekordów bibliografii załącznikowych, identyfikatorów oraz form wydawniczych),
- poprzez pobranie całego korpusu jako archiwum zawierające zarówno dane, jak i oprogramowanie umożliwiające uruchomienie całego systemu w ramach własnej infrastruktury sprzętowej,
 - poprzez pobranie jednym z wybranych formatów (RDF, tabela, arkusz Google, tabela przestawna) jedynie wycinka danych spełniającego kryteria zapytania wysłanego do systemu (<http://opencitations.net/sparql>), a napisanego w języku SPARQL¹⁴ (jest to język zapytań dedykowany przetwarzaniu danych o strukturze sieciowej, a fragment przykładu wyrażenia w nim zapisanego przedstawiono na rys.2.),
 - poprzez wysłanie zapytania SPARQL wprost do procesu nasłuchowego serwera z pominięciem graficznego interfejsu użytkownika.



Rys. 2. Przykład zapytania w języku SPARQL dla platformy OpenCitations.
Źródło: OCC SPARQL endpoint GUI. W: *OpenCitations* [online]. [Dostęp 16.08.2017].
Dostępny w: <http://opencitations.net/sparql>.

Autorzy projektu, mając świadomość jego obecnych ograniczeń funkcjonalnych polegających głównie na ubogich możliwościach analizowania danych wprost z poziomu udostępnionej platformy, informują, że w chwili obecnej użytkownik chcący przeglądać całą zawar-

¹⁴ SPARQL Query Language for RDF [online]. W3C, 2017. [Dostęp 16.08.2017]. Dostępny w: <https://www.w3.org/TR/rdf-sparql-query/>.

tość korpusu skazany jest właściwie jedynie na samodzielne konstruowanie zapytań w języku SPARQL lub bardzo prosty interfejs graficzny, który w dodatku udostępnia jedynie dane o jednostkach bibliograficznych. Zapowiadają równocześnie rychłą rozbudowę tych możliwości. Obietnice te wydają się mieć pokrycie w informacjach o rozbudowie projektu¹⁵ ze środków przyznanych przez fundację Alfreda P. Sloana¹⁶.

Podsumowanie

Choć początki inicjatywy OpenCitations sięgają już roku 2010, to dopiero modernizacje przeprowadzone w ciągu ostatnich dwóch lat sprawiają, że wyróżnia się ona na tle wielu innych rozwiązań. Oparcie modelu informacyjnego na udokumentowanych standardach sprawia, że tak zbudowany indeks cytowań można łatwo zasilać informacjami z nowych źródeł, a dane z niego pobierane mogą być interpretowane w sposób jednoznaczny. Toczące się po dzień dzisiejszy dyskusje czy inicjatywy krajowych bibliograficznych baz danych powinny być scentralizowane, czy też raczej rozproszone dziedzinowo¹⁷, byłyby drugoplanowe, gdyby konstruując takie bazy, przyjąć wspólny referencyjny model składowanych w nich danych. Dane składowane w ten sposób i opatrywane globalnymi dla całej inicjatywy identyfikatorami mogą stanowić w zasadzie takie samo źródło analiz czy będą składowane w postaci scentralizowanej czy rozproszonej (przykładowo w język SPARQLwbudowane są już możliwości odpytywania wielu źródeł jednocześnie). Niestety, krajowe inicjatywy borykają się również z bardziej podstawowymi problemami¹⁸, jak chociażby powielone opisy bibliograficznych jednostek cytowanych¹⁹ czy też brak możliwości pozyskania danych w jakiegokolwiek postaci strukturalnej umożliwiającej dalsze analizy – dane udostępniane są zwykle jedynie za pomocą interfejsów graficznych użytkownika, czyli stron HTML.

Biorąc pod uwagę powyższe, choć trudno przewidywać, kiedy opisywana inicjatywa okaże się bardziej interesująca dla naukowców od komercyjnych indeksów cytowań, to z całą pewnością warto zapoznać się z podejściem do jej budowy (abstrahując nawet od konkretnych wybranych standardów czy silnika zarządzania bazą danych) podczas planowania modernizacji podobnych platform w systemach krajowych.

Bibliografia:

1. *Alfred P. Sloan Foundation* [online]. [Dostęp 16.08.2017]. Dostępny w: <https://sloan.org/>.
2. BEDNAREK-MICHALSKA, B. Bibliograficzne bazy danych : perspektywy i problemy rozwoju. Sprawozdanie z konferencji w Krakowie, 26–27 czerwca 2017 r. *Biuletyn EBIB* [online]. 2017, No 173.

¹⁵ SHOTTON, D. *The Sloan Foundation funds OpenCitations* [online]. OpenCitations, 2017. [Dostęp 16.08.2017]. Dostępny w: <https://opencitations.wordpress.com/>.

¹⁶ *Alfred P. Sloan Foundation* [online]. [Dostęp 16.08.2017]. Dostępny w: <https://sloan.org/>.

¹⁷ BEDNAREK-MICHALSKA, B. Bibliograficzne bazy danych: perspektywy i problemy rozwoju. Sprawozdanie z konferencji w Krakowie, 26–27 czerwca 2017 r. *Biuletyn EBIB* [online]. 2017, No 173. [Dostęp 16.08.2017]. ISSN 1507-7187. Dostępny w: <http://open.ebib.pl/ojs/index.php/ebib/article/view/544>.

¹⁸ KAMIŃSKA, A.M. ProBIT – prospektywna metoda tworzenia..., *dz. cyt.* (w druku).

¹⁹ Przykłady oraz sposoby identyfikowania i przeciwdziałania powstawaniu takich anomalii znaleźć można w: KAMIŃSKA, A.M. Miary podobieństw łańcuchów znakowych a deduplikacja rekordów w bibliograficznych bazach danych. *Przełęcz Biblioteczny* 2017, nr 4 (w druku).

- [Dostęp 16.08.2017]. ISSN 1507-7187. Dostępny w:
<http://open.ebib.pl/ojs/index.php/ebib/article/view/544>.
3. *Blazegraph* [online]. [Dostęp 16.08.2017]. Dostępny w: <https://www.blazegraph.com/>.
 4. *Europe PMC* [online]. [Dostęp 16.08.2017]. Dostępny w: <http://europepmc.org/>.
 5. GARFIELD, E. Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science* 1955, Vol. 122, No. 3159, s. 108–111. ISSN 0036-8075.
 6. *Graffo: Graphical Framework for OWL Ontologies* [online]. [Dostęp 16.08.2017]. Dostępny w: <http://www.essepuntato.it/graffoo/>.
 7. *Initiative for Open Citations* [online]. [Dostęp 16.08.2017]. Dostępny w: <https://i4oc.org>.
 8. KAMIŃSKA, A.M. Miary podobieństw łańcuchów znakowych a deduplikacja rekordów w bibliograficznych bazach danych. *Przegląd Biblioteczny* 2017, nr 4 (w druku).
 9. KAMIŃSKA, A.M. O rozwoju graficznych języków komunikacji. *Zagadnienia Informacji Naukowej* 2017, nr 2 (w druku).
 10. KAMIŃSKA, A.M. Od druków źródłowych po mapy nauki. Bibliograficzna baza danych GRUBA. W: KOWALSKA, M., OSIŃSKA, V. (red.). *Wizualizacja informacji w humanistyce*. Toruń: Wydaw. Uniwersytetu Mikołaja Kopernika, 2017 (w druku).
 11. KAMIŃSKA, A.M. ProBIT – prospektywna metoda tworzenia trawersowalnych indeksów cytowań a współczesne problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych. *Zagadnienia Informacji Naukowej* 2017, nr 1, s. 66–82.
 12. KAMIŃSKA, A.M. Tam, gdzie zaczyna się bibliometria, czyli jak pozyskać materiał analityczny z autopsji. *Biuletyn EBIB* [online]. 2017, No 173. [Dostęp 16.08.2017]. ISSN 1507-7187. Dostępny w: <http://open.ebib.pl/ojs/index.php/ebib/article/view/534>.
 13. *OCO, the OpenCitations Ontology* [online]. OpenCitations, 2016. [Dostęp 16.08.2017]. Dostępny w: <http://opencitations.net/ontology.html>
 14. PERONI, S., SHOTTON, D. Metadata for the OpenCitations Corpus, 2017. [Dostęp 16.08.2017]. Dostępny w: <https://dx.doi.org/10.6084/m9.figshare.3443876>.
 15. PERONI, S., SHOTTON, D. One year of the OpenCitations Corpus: Releasing RDF-based scholarly citation data into the Public Domain. W: *Proceedings of the 16th International Semantic Web Conference (ISWC 2017)* [online]. [Dostęp 16.08.2017]. Dostępny w: <https://w3id.org/people/essepuntato/papers/oc-iswc2017.html>.
 16. *RDF* [online]. W3C, 2017. [Dostęp 16.08.2017]. Dostępny w: <https://www.w3.org/RDF/>.
 17. SHOTTON, D. *The Sloan Foundation funds OpenCitations* [online]. OpenCitations, 2017. [Dostęp 16.08.2017]. Dostępny w: <https://opencitations.wordpress.com/>.
 18. *SPARQL Query Language for RDF* [online]. W3C, 2017. [Dostęp 16.08.2017]. Dostępny w: <https://www.w3.org/TR/rdf-sparql-query/>.