

## Identyfikacja autorów w serwisach dla naukowców na przykładzie ORCID

### Streszczenie

Rosnące wykorzystanie baz danych, repozytoriów oraz serwisów społecznościowych do ilościowych analiz naukometrycznych spowodowało potrzebę wypracowania precyzyjnych mechanizmów identyfikacji autora i jego publikacji. Szczególnie ważne jest to dla osób, które stosowały różne formy zapisu nazwiska i dla autorów o takim samym imieniu i nazwisku publikujących w podobnych dziedzinach. Odpowiedzią na te problemy są powstające systemy, które identyfikują autorów nie tylko poprzez imię i nazwisko, ale dzięki nadawaniu im specjalnych kodów. W artykule przedstawiono wybrane systemy identyfikacji autorów ze szczególnym omówieniem najnowszego narzędzia ORCID (Open Researcher and Contributor ID).

### Abstract

Increasingly common use of databases, repositories and social services for the purpose of quantitative scientometric analyses has generated the need for accurate mechanisms of identification of an author and publications. It is of particular significance for non-English surnames, for the people who use various forms of spelling of their surname as well as for several authors with the same name and surname. The created systems of authors identification not only via their name and surname, but also via giving them special codes constitute solution to these problems. The paper presents selected systems of authors identification with particular emphasis on the most recent tool - ORCID (Open Researcher and Contributor ID).

### 1. Wstęp

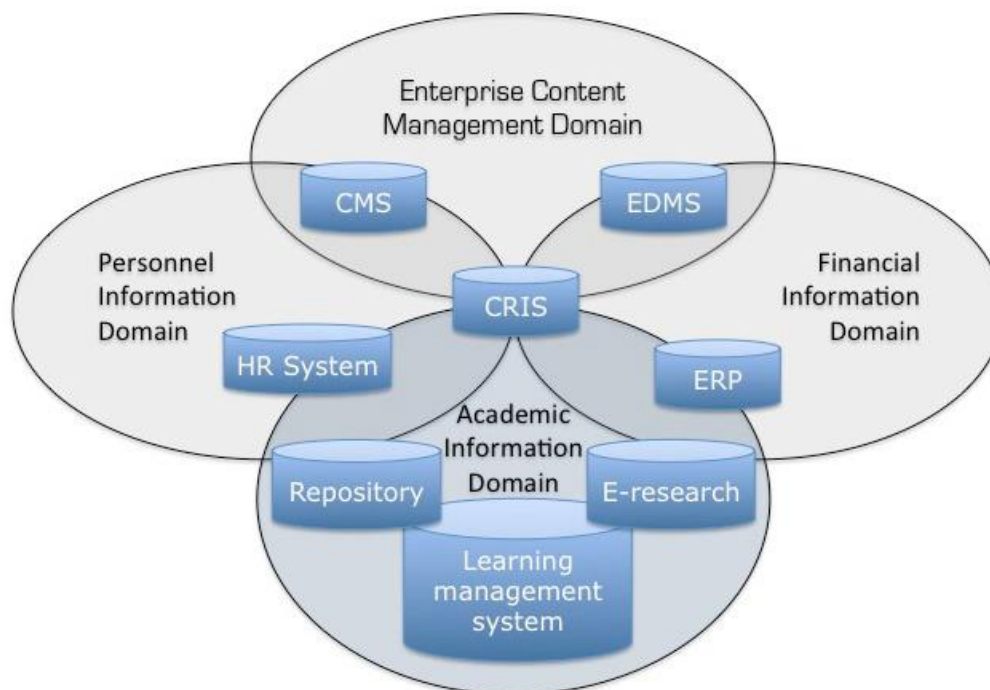
Systemy informatyczne wykorzystywane w środowisku naukowym w obszarze informacji naukowej stanowią komputerową reprezentację jakiegoś fragmentu świata rzeczywistego związanego z różnymi aspektami działalności badawczej naukowców. Budowane są poprzez modelowanie polegające na odwzorowaniu rzeczywistych obiektów w bazie danych i utworzeniu pomiędzy nimi mechanizmu powiązań<sup>1</sup>. W obszarze związanym z badaniami naukowymi informatyzacją można objąć obiekty materialne takie jak instytucja, pracownik, publikacja, oraz niematerialne takie jak wiedza, metodyka czy technologia badań. Informacje o obiektach są opisywane za pomocą jednakowych cech (atrybutów) i przechowywane w bazie. Atrybuty te powinny jednoznacznie identyfikować obiekt - dane o autorze są

---

<sup>1</sup> Brigitte Jörg, Thorsten Höllrigl, Sicilia Miguel-Angel: Entities and Identities in Research Information Systems. In: Keith G Jeffery, Jan Dvořák (red.): E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production, 2012, s. 185-194. Dostęp: [http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012\\_20\\_full\\_paper.pdf](http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012_20_full_paper.pdf) [odczyt: 23.07.2013]

zazwyczaj opisane poprzez imię i nazwisko, natomiast dane dotyczące publikacji reprezentowane są poprzez opis bibliograficzny zawierający określone elementy (tytuł, nazwę czasopisma, rok wydania, strony). Jeśli chodzi o zasady powiązań to tworzone są według reguł, które pozwalają na identyfikację i zebranie informacji o obiekcie (np. autorze) wewnątrz systemu. Identyfikacja autora występuje zarówno w bazach dokumentujących publikacje, jak też w systemach rejestrujących projekty badawcze, czy w systemach kadrowych rejestrujących pracowników danej instytucji. Obecnie środowisko nauki przyjmuje zasięg międzynarodowy, dlatego też powstaje potrzeba tworzenia jednoznacznych sposobów identyfikowania obiektów w różnych systemach i bazach danych na szeroką skalę. Uniwersalny mechanizm identyfikacji pracowników nauki, publikacji, instytucji badawczych, projektów naukowych staje się potrzebny w celu zapewnienia szybkiej, a zarazem precyzyjnej nawigacji w globalnych zasobach wiedzy.

Przykładowym systemem baz danych, przechowującym informacje o prowadzonych projektach badawczych, instytucjach i osobach związanych z badaniami oraz o programach finansowania badań w Europie jest CRIS – Current Research Information System<sup>2</sup>. Pełni on rolę centrum informacji o środowisku badawczym dla różnych zastosowań w zakresie zarządzania badaniami naukowym w Europie. Do sprawnego działania zintegrowanych modułów (rys. 1) zastosowano w nim uniwersalne identyfikatory obiektów.



Rys. 1. Poszerzony model Academic Information Domain<sup>3</sup>

<sup>2</sup> euroCRIS – Current Research Information System. Dostęp: <http://www.eurocris.org/> [odczyt: 18.07.2013]

<sup>3</sup> Karen van Godtsenhoven, Mikael Karstensen Elbaek, Barbara Sierman [et al.]: Emerging Standards for Enhanced Publications and Repository Technology: Survey on Technology. Amsterdam Univ. Press, 2009. Dostęp: <http://dare.uva.nl/document/150752> [odczyt: 22.08.2013]

Celem CRISu jest ułatwienie dokumentowania, raportowania i zarządzania procesami badań dla instytucji finansujących w zakresie alokacji środków, oceny projektów, rozliczania i rozpowszechniania głównych wyników badań. Dla instytucji badawczych jest to narzędzie wspomagające planowanie, śledzenie stanu realizacji i rozliczania projektów. Dla indywidualnych użytkowników CRIS jest źródłem informacji o funduszach, pozwala uniknąć powielania badań, ułatwia śledzenie pełnych tekstów publikacji naukowych, nawiązywanie kontaktów pomiędzy badaczami. Daje możliwość generowania naukowego CV, zarządzania informacją o projektach, tworzenia bibliografii i raportów torując drogę do wdrożeń i komercjalizacji badań.

## **2. Unikalne identyfikatory w systemach informacji naukowej**

Globalizacja informacji powoduje, że każdy obiekt informacyjny powinien być dokładnie zdefiniowany i jednoznacznie nazwany. Aby uniknąć przekłamań stosuje się przypisywanie im unikalnych identyfikatorów umożliwiających przetwarzanie danych w sposób odpowiedni do ich znaczenia. Ułatwieniem dla środowiska naukowego będzie więc stosowanie dla autorów, publikacji, instytucji, patentów czy projektów, standaryzowanych numerów identyfikacyjnych zarejestrowanych i nadzorowanych przez odpowiednie instytucje.

Przykładami międzynarodowych systemów służących wystandaryzowanej identyfikacji obiektów związanych ze sferą badawczą i publikacyjną są:

ISO 27729: International Standard Name Identifier (ISNI)

ISO 26324: Digital Object Identifier System (DOI)

ISO 21047: International Standard Text Code (ISTC)

Standard ISNI – International Standard Name Identifier stosowany od 2012 roku, jest narzędziem służącym kontroli danych dotyczących osób i instytucji zaangażowanych w działalność twórczą (<http://www.insni.org>). Pozwala rozpoznać autorów i śledzić ich dorobek w dużych zbiorach dokumentów cyfrowych poprzez przydzielone identyfikatory. Ułatwia to wyszukiwanie autorów w przypadkach kiedy stosują różne formy zapisu nazwiska czy imienia lub publikują pod zmienionym nazwiskiem. ISNI składa się z 16 znaków, pozwala śledzić dorobek twórczy w różnych sferach działalności (np. nauka, literatura, muzyka). Obecnie działają dwie agencje rejestracyjne (Bowker, Ringgold), które przydzieliły już ponad 6,5 mln identyfikatorów, w tym dla 6,4 mln osób i 100 tys. instytucji<sup>4</sup>.

System DOI – Digital Object Identifier (<http://www.doi.org/>) to cyfrowy identyfikator dokumentu elektronicznego, który nie zależy od fizycznej lokalizacji dokumentu, lecz jest do niego na stałe przypisany. Dokument zachowuje swój identyfikator DOI przez cały okres istnienia, a jeżeli zostanie kiedykolwiek usunięty, jego numer nie może być ponownie wykorzystany. DOI może być utworzony dla każdej jednostki cyfrowej, tj.: rozdziału z książki,

---

<sup>4</sup> Agnieszka Włudyka-Szeliga: Do czego ISTC? Nowe narzędzie przetwarzania danych w przemyśle książkowym. Biblioteka Analiz nr 22/2012 (numer specjalny), s. 25-32

artykułu, numeru czasopisma, wykresu, oprogramowania czy utworu muzycznego. Zapewnia możliwość zarządzania prawami wydawniczymi, usprawnia identyfikację, zakup lub dostawę dokumentu<sup>5</sup>.

Standard ISTC – International Standard Text Code określa każdy indywidualny format obiektu tekstowego (<http://www.istc-international.org>), funkcjonuje od 2008 roku. Identyfikuje zarówno książkę w postaci fizycznej jak i wersję cyfrową. Stosuje się go do wszystkich edycji i wariantów tego samego tekstu, identyfikując i grupując w ten sposób wszystkie formaty podstawowego tekstu<sup>4</sup>. ISTC składa się z 16 znaków, funkcjonuje w połączeniu z bazami danych i wyszukiwarkami publikacji. Agencje rejestrujące ISTC: Soci t  de gestion de la BTLF z Kanady, Beijing Chinese ISTC Agency Ltd., CB z Holandii, Cercle de la Librairie-Electre z Francji, Meta4Books z Belgii, MVB z Niemiec, Nielsen z Wielkiej Brytanii, Australii i Nowej Zelandii, FIPR z Rosji i polskie Stowarzyszenie Autor w i Wydawc w "Polska Ksi żka".

### 3. Systemy identyfikacji obiekt w

Systemy identyfikacji os b w  rodowisku naukowym s  cz sto wykorzystywane w instytucji jako identyfikator pracownika, w bibliotece naukowej jako identyfikator czytelnika, w bazie dorobku publikacyjnego jako identyfikator autora. Mog  te  by  stosowane jako numer PBN ID dla autora w Polskiej Bibliografii Naukowej (POL-on), a w bazach mi dzynarodowych jako ResearcherID (WoS) czy Author ID (Scopus). S  te  inne inicjatywy np. DAI – Digital Author Identifier w Holandii, system Author Service rejestruj cy autor w z dziedziny ekonomii, AuthorClaim w bazach Academic Contributor Information System czy otwarty standard OpenID zapewniaj cy automatyczne logowanie w r wnych serwisach internetowych<sup>6</sup>.

Uniwersalnym standardem wykorzystywanym globalnie staje si  wspomniany wcze niej mi dzynarodowy identyfikator tw rc w INSI oraz coraz cz ciej rozpowszechniany Open Researcher and Contributor ID (ORCID) dla  rodowiska naukowego.

W obszarze nauki istotna jest te  identyfikacja publikacji. Nadawanie identyfikator w dla prac odbywa si  w bazach danych poprzez nadanie unikalnego numeru dla zarejestrowanego dokumentu. W serwisie PubMed jest to numer PMID, w Web of Science odpowiada za to Accession Number. Natomiast do zlokalizowania pe nego tekstu dokumentu elektronicznego najcz ciej wykorzystuje si  wspomniany wcze niej numer DOI.

Jednostki organizacyjne to kolejny obiekt wykorzystywany na wielu obszarach w systemach informacji w  rodowisku badawczym. W bazie Scopus nadawany jest Affiliation ID, w systemie POL-on przypisywany jest PBN ID dla jednostki organizacyjnej, w Komisji Europejskiej w wszystkie instytucje uczestnicz ce w programach ramowych maj  przydzielony Participant Identification Code (PIC) wykorzystywany przy sk adaniu, zarz dzaniu i raportowaniu projekt w.

---

<sup>5</sup> Justyna Maczuga, Jolanta Przytu ska: Czy publikacje naukowe powinny mie  DOI? W: Standardy w archiwistyce cyfrowej. Oficyna Wydawnicza AFM 2013 (w druku)

<sup>6</sup> El zbieta Jankowska: ResearcherID i inne systemy identyfikacji autora. Forum Uczelniane, 1(9) 2011, s. 24-25

#### 4. ORCID jako identyfikator autora w globalnej sieci naukowej

System identyfikacji autorów ORCID to międzynarodowa, otwarta inicjatywa stworzona w 2010 roku, która ma na celu opracowanie i utrzymanie rejestru unikalnych identyfikatorów dla naukowców bez względu na dyscyplinę nauki czy granice państw. System zapewnia:

- unikalny rejestr identyfikatorów dla autorów,
- pomoc w obsłudze i uwierzytelnianiu danych.

ORCID jest bezpłatny i mogą z niego korzystać:

- osoby indywidualne, które mogą zarządzać swoim kontem, uzupełniać je i przeszukiwać pozostałe rekordy,
- instytucje, które mogą aktualizować dane i rejestrować swoich pracowników i studentów.


ORCID został utworzony po to, aby pomóc w rozwiązaniu problemu identyfikacji autorów badań naukowych w przypadku bardzo popularnych, powtarzających się nazwisk, identyfikacji publikacji osób piszących artykuły z tej samej dziedziny, powiązania w jeden rekord prac jednego autora bez względu na zmianę afiliacji podyktowaną np. zmianą pracy. ORCID zapewnia komunikację naukową poprzez utworzenie centralnego rejestru unikalnych identyfikatorów dla poszczególnych badaczy. Przestrzega zasad prywatności stosując mechanizmy zabezpieczenia i ochrony danych wprowadzonych do systemu.

ORCID tworzą przedstawiciele środowisk naukowych (Harvard University, Cornell University) i wydawców (Wiley-Blackwell, Thomson Reuters, Elsevier, Nature) wspierani przez instytucje IT jak: California Institute of Technology, National Institute of Informatics czy Association for Computing Machinery.

Numer identyfikacyjny ORCID składa się z 16 cyfr podzielonych na 4 części zapisanych w formie HTTP URI (np. <http://orcid.org/0000-0002-4599-1309>). Nie zawiera on informacji semantycznych takich jak rok rejestracji użytkownika (jak w przypadku np. Researcher ID) czy kod kraju z którego pochodzi.

Identyfikator ma możliwość łączenia i linkowania do profili autora założonych w innych, podobnych systemach identyfikujących (Scopus, Researcher ID) tworząc spójną całość.

Aby zarejestrować się w bazie ORCID i otrzymać swój unikalny numer należy wejść na stronę <https://orcid.org/register>. W formularzu rejestracyjnym wpisuje się podstawowe dane tj. imię, nazwisko, e-mail, hasło. Należy również określić czy dostęp do publikacji, które wprowadzimy będzie całkowicie otwarty, półotwarty (dostęp tylko dla osób, które sami wskażemy) lub ograniczony tylko do właściciela konta. Rys. 2 pokazuje uzupełniony profil naukowca uwzględniający informacje takie jak: biografia, słowa kluczowe, strona internetowa autora czy instytucji, identyfikatory w innych systemach czy lista publikacji.



Connecting Research  
and Researchers

FOR RESEARCHERS

FOR ORGANIZATIONS

ABOUT

HELP

SIGN IN

English

[SIGN IN](#)
[REGISTER FOR AN ORCID ID](#)

352041 ORCID iDs and counting. [See more...](#)

**Jolanta Przuluska**  
<http://orcid.org/0000-0002-4599-1309>

**Keywords:** bibliometrics, research evaluation, research policy, research management

**Websites:**  
[Nofer Institute of Occupational Medicine](#)

**Other IDs:**  
 ResearcherID: E-9440-2010  
 Scopus Author ID: 6602817747

**Personal Information**

**Biography**

Education: Faculty of Chemistry, Technical University, Lodz, Poland; doctorate in medicine: Nofer Institute of Occupational Medicine, Lodz, Poland;

Postgraduate education: Scientific Information and Librarianship, Jagiellonian University, Cracow, Poland;

Employment: Department of Knowledge Management, Nofer Institute of Occupational Medicine, Lodz (1989 to present). Head of Scientific Library (since 2000) and Scientific Information Section (since 2011). Lecturer at postgraduate courses on the use of bibliographic data and scientific publications and medical retrieval, for physicians and employees of sanitary and epidemiological stations.

Main fields of interest: Mainly interested in bibliometrics and information systems in toxicology, occupational medicine, environmental and public health.

---

**Publications**

**HOW TO INCREASE THE VISIBILITY OF YOUR SCIENTIFIC PUBLICATION ON THE INTERNET?:** [Medycyna Pracy 2012](#)

Maczuga, Justyna and Przuluska, Jolanta, (2012). "HOW TO INCREASE THE VISIBILITY OF YOUR SCIENTIFIC PUBLICATION ON THE INTERNET?", *Medycyna Pracy*, vol. 63, no. 4, pp. 391-393

Rys. 2. Profil naukowca w serwisie ORCID

## 5. Podsumowanie

Idea tworzenia globalnych, unikalnych identyfikatorów dla środowiska naukowego jest stopniowo wdrażana w życie i zapewne będzie ułatwieniem dla wielu środowisk zarządzających wiedzą. Integracja systemów informatycznych, wykorzystanie stałych identyfikatorów obiektów pozwoli na przypisanie badań do osoby, zapewni precyzyjną identyfikację naukowców z obszarami ich aktywności zawodowej, niezależnie od instytucji i kraju w jakim obecnie pracują. Unikalny identyfikator używany przez cały okres aktywności zawodowej ułatwi dokumentowanie i raportowanie badań oraz usprawni procesy zarządzania projektami naukowymi.

## Bibliografia

euroCRIS - Current Research Information System. Dostęp: <http://www.eurocris.org/> [odczyt: 18.07.2013]

Godtsenhoven Karen van, Elbaek Mikael Karstensen, Sierman Barbara [et al.]: *Emerging Standards for Enhanced Publications and Repository Technology: Survey on Technology*. Amsterdam Univ. Press, 2009. Dostęp: <http://dare.uva.nl/document/150752> [odczyt: 22.08.2013]

Jankowska Elżbieta: ResearcherID i inne systemy identyfikacji autora. *Forum Uczelniane*, 1(9) 2011, s. 24-25

Jörg Brigitte, Höllrigl Thorsten, Miguel-Angel Sicilia: Entities and Identities in Research Information Systems. In: Keith G Jeffery, Jan Dvořák (red.): E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production, 2012, s. 185-194. Dostęp: [http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012\\_20\\_full\\_paper.pdf](http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012_20_full_paper.pdf) [odczyt: 23.07.2013]

Maczuga Justyna, Przyłuska Jolanta: Czy publikacje naukowe powinny mieć DOI? W: Standardy w archiwistyce cyfrowej. Oficyna Wydawnicza AFM 2013 (w druku)

Włudyka-Szeliga Agnieszka: Do czego ISTC? Nowe narzędzie przetwarzania danych w przemyśle książkowym. Biblioteka Analiz nr 22/2012 (numer specjalny), s. 25-32