



Biblioteka **Główna**  
Politechniki **Warszawskiej**

# Otwarte dane badawcze - kolejny krok do otwierania nauki

Bibliograficzne bazy danych : perspektywy i problemy rozwoju  
III Konferencja Naukowa Konsorcjum BazTech  
Kraków, 26-27.06.2017

Edyta Strzelczyk

[e.strzelczyk@bg.pw.edu.pl](mailto:e.strzelczyk@bg.pw.edu.pl)





# Dane badawcze

## definicje

Zarejestrowane materiały o charakterze faktograficznym, powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych.

*Open Research data. Definitions*

Informacja, w szczególności zebrane fakty, liczby, które mogą posłużyć badaniom i być traktowane jako podstawa do dalszego wnioskowania, dyskusji lub obliczeń.

*Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*





# Dane badawcze

## definicje

Zapisy faktów będących rezultatem badań naukowych (np. obserwacji, pomiarów, doświadczeń, eksperymentów, itp.), na których to opiera się wnioskowanie naukowe (...) mogą to być surowe dane (*raw data*) lub takie, które zostały już w jakiś sposób oczyszczone, uporządkowane czy też w inny sposób przetworzone, zapisane w dowolnym formacie i na dowolnym nośniku.

Queensland University of Technology

Zaliczane są także do nich informacje na temat pochodzenia tych danych: jak, gdzie, kiedy zostały zebrane oraz za pomocą jakich narzędzi, czyli obejmują całą dokumentacja badań (...). W skład danych badawczych wchodzi także metadane (np. nazwa zbioru danych, autorzy, itp.) stanowiące podstawowy opis umożliwiający identyfikację i zrozumienie danych.

The University of Melbourne





# Przykładowe dane badawcze

- dokumenty tekstowe, notatki,
- dane liczbowe (w formie nieprzetworzonej lub czytelnej dla komputera),
- kwestionariusze ankiety, wyniki badań ankietowych,
- nagrania audio i video, zdjęcia,
- modele matematyczne, algorytmy, schematy,
- oprogramowanie (skrypty, pliki wejściowe),
- wyniki symulacji komputerowych,
- protokoły laboratoryjne, opisy metodologiczne,
- próbki, artefakty, obiekty,
- metadane (opisy)





## Otwarte dane badawcze - tzn. jakie?

(...) takie, które są dostępne w postaci cyfrowej w Internecie bez żadnych ograniczeń, czyli mogą być dowolnie używane, rozpowszechniane i przetwarzane przez kogokolwiek, gdziekolwiek w dowolnym celu.



Można się dowiedzieć, że istnieją



Można je obejrzeć



Można z nich skorzystać





# Ograniczenia

## finansowe

---

- wiedza jako towar rynkowy
- opłaty za dostęp do danych

## techniczne

---

- zapis w formacie utrudniającym odczytanie
- specjalistyczne oprogramowanie
- formaty uniemożliwiające ponowne wykorzystanie danych

## prawne

---

- prawo autorskie
- prawo osób trzecich
- prawa pokrewne
- prawo przysługujące pracodawcy
- tzw. *sui generis* do baz danych





# Przeszkody / obawy



uregulowania  
prawne



standardy



system oceny



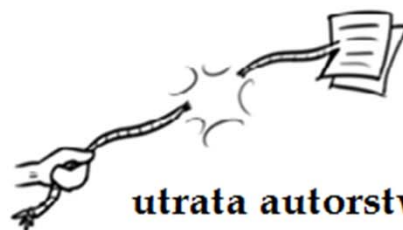
prywatność



utrata dochodu



utrata wartości  
marki



utrata autorstwa



utrata kontroli



utrata  
potencjalnego  
dochodu



efektywniejsze  
wykorzystanie



utrata jakości



utrata spójności



obawa przed  
błędami



dodatkowa praca





# Otwarte dane badawcze

-  Wytworzone i zapisane
-  Przechowywane
-  Dostępne
-  Łatwe do znalezienia
-  Zrozumiałe
-  Zaufane
-  Interoperacyjne
-  Odtwarzalne







## Open Research Data Pilot (ORD Pilot)

- **Wymagania:** udostępnienie w domenie publicznej danych badawczych z badań finansowanych w ramach programu Horyzont 2020
- **Przedmiot:** Dane wraz z metadanymi niezbędne do weryfikacji wyników badań zaprezentowanych w publikacji naukowej oraz inne dane wybrane przez naukowców
- **Czas:** tak szybko jak to możliwe, nie później niż 6 miesięcy po publikacji wyników
- **Wyłączenie:**
  - komercyjne lub przemysłowe wykorzystanie danych,
  - wymogi poufności związane z bezpieczeństwem,
  - ochrona danych osobowych,
  - niemożliwość osiągnięcia głównego celu,
  - brak danych.





## Wymagania ORD Pilot

- Zdeponowanie danych w repozytorium
- Umożliwienie dostępu do danych badawczych, ich maszynową analizę, ponowne wykorzystanie, kopiowanie i rozpowszechnienie bez opłat ze strony użytkownika
- Udostępnienie informacji o narzędziach niezbędnych do weryfikacji wyników lub dostarczenie wymaganych narzędzi
- Opracowanie planu zarządzania danymi (*Data Management Plan - DMP*)





# Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce

październik 2015

*„...zaleca, aby krajowe podmioty finansujące badania naukowe ze środków publicznych (...) stosowały i upowszechniały zasady, zgodnie z którymi publikacje i dane badawcze powstające w wyniku finansowanych lub współfinansowanych przez nie badań znajdą się w otwartym dostępie.”*

*„obejmujących przede wszystkim selekcjonowanie do publikacji tylko takich danych badawczych lub tylko tak opracowanych, które nie spowodują naruszenia tajemnic ani interesów związanych z komercjalizacją. Docelowo powszechny dostęp do wyników badań naukowych (zarówno publikacji, jak i danych badawczych) (...) powinien być ogólną zasadą (...)”*



Ministerstwo Nauki  
i Szkolnictwa Wyższego





# Repozytoria danych badawczych



re3data.org Reaches a  
Milestone and Begins

Enhancements to  
creating and updating

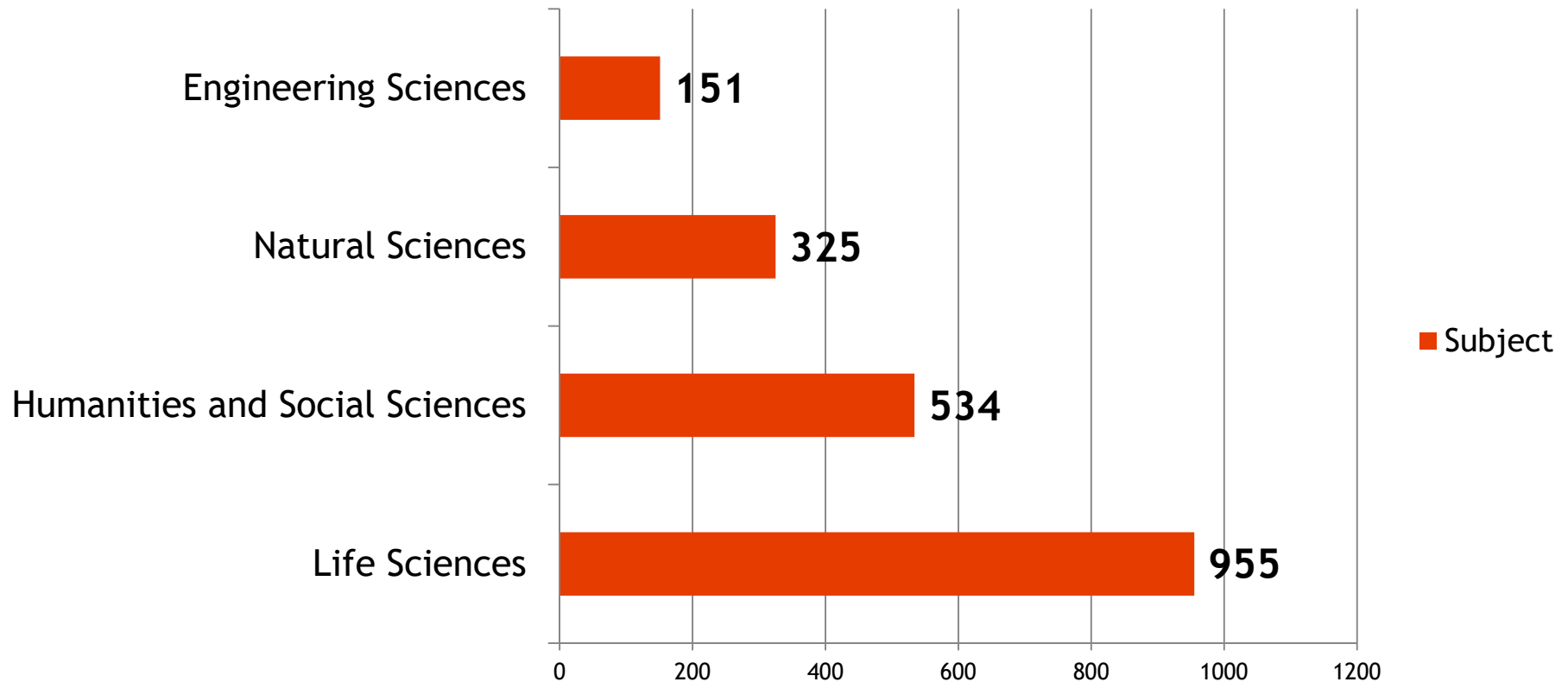
New re3data.org  
Schema and Search





# Repozytoria danych badawczych

## Browse by subject



Źródło: opracowanie własne na podstawie: <http://www.re3data.org/>





# Sytuacja w Polsce - w świetle badań

## Zapotrzebowanie

	Rozprawy doktorskie	Publikacje naukowe	Informacje o posiadanych patentach	Informacje o produktach (technologie, licencje, materiały, oprogramowanie, itp.)	Informacje o posiadanych urządzeniach/stanowiskach wyspecjalistycznych	Informacje o laboratoriach udostępnionych do wykorzystania komercyjnego	Informacje o obiektach muzealnych	Informacje o zrealizowanych projektach	Wyniki badań naukowych – dane badawcze	Informacje o pracownikach naukowych i ich dorobku	Informacje o ekspertach i obszarze ich ekspertyzy
naukowcy	7,48	9,50	6,02	6,36	6,45	6,38	4,37	6,74	8,44	7,48	7,43
przedsiębiorcy	5,8	8,25	7,2	7,35	7	7,85	3,25	6,3	7,6	6,3	7,1
pracownicy admin. publicznej	4,45	6,59	3,77	5,91	4,68	4,05	4,59	5,68	5,82	5,00	6,18
doktoranci	7,74	9,37	6,00	6,37	6,68	6,32	3,84	5,63	7,95	7,32	7,53
studenci	6,49	8,89	6,47	7,13	6,47	6,27	4,20	6,00	8,33	6,89	6,93
Ogólnie	6,75	8,82	5,80	6,41	6,30	6,20	4,34	6,43	8,02	7,00	7,13

Poziom zainteresowania poszczególnych grup odbiorców otwartym dostępem do różnego typu danych  
 Źródło: opracowanie własne

N = 424  
 naukowcy = 65%  
 pracownicy admin.  
 publicznej = 7%

przedsiębiorcy = 7%  
 doktoranci = 6%  
 studenci = 15%





# Sytuacja w Polsce - w świetle badań

## Zapotrzebowanie

	Interesują mnie dane w postaci surowej.	Interesują mnie dane w postaci opracowanej (interpretacje oraz wnioski).	Interesują mnie dane udostępniane nieodpłatnie.	Jestem gotowy zapłacić za dostęp do danych.	Interesują mnie narzędzia analityczne na poziomie systemu udostępniającego dane.	Możliwość pobrania danych na własny komputer jest dla mnie istotna.	Przeglądanie danych online bez możliwości ich pobrania na własny komputer jest całkowicie wystarczające.	Jestem zainteresowany dostępem do narzędzi wizualizujących dane.	Jestem zainteresowany zaawansowanymi możliwościami wielokryterialnego wyszukiwania obiektów.	Jestem zainteresowany wyszukiwaniem obejmującym jednocześnie różne instytucje i obszary specjalizacji.	Wysoka jakość wizualizacji danych jest dla mnie ważna.	Udostępnianie danych w popularnych formatach (PDF, XML) jest bardzo ważne.	Format udostępnianych danych jest bez znaczenia.	To bardzo ważne, aby publiczne jednostki naukowe nieodpłatnie udostępniały wyniki swoich prac.
naukowcy	6,51	8,48	9,54	3,81	6,34	8,52	4,16	7,10	7,87	7,93	7,10	8,76	3,15	8,47
przedsiębiorcy	6,40	8,65	9,40	5,50	6,05	8,70	4,70	6,70	7,50	7,60	6,35	8,40	3,80	8,05
pracownicy admin. publicznej	3,41	7,09	8,59	2,14	4,68	7,09	4,95	5,86	6,41	6,50	5,50	7,41	2,36	7,23
doktoranci	7,74	8,68	9,21	3,74	5,79	8,58	4,05	6,95	7,58	7,53	6,16	8,32	3,58	7,42
studenci	6,82	8,13	9,40	3,71	6,62	8,38	4,16	7,24	7,73	7,91	6,53	8,49	2,98	8,22
Ogólnie	6,31	8,25	9,30	3,83	6,15	8,36	4,31	6,92	7,63	7,73	6,75	8,48	3,19	8,13

Poziom zgodności grup respondentów z określonymi aspektami związanymi z OD.

Źródło: opracowanie własne

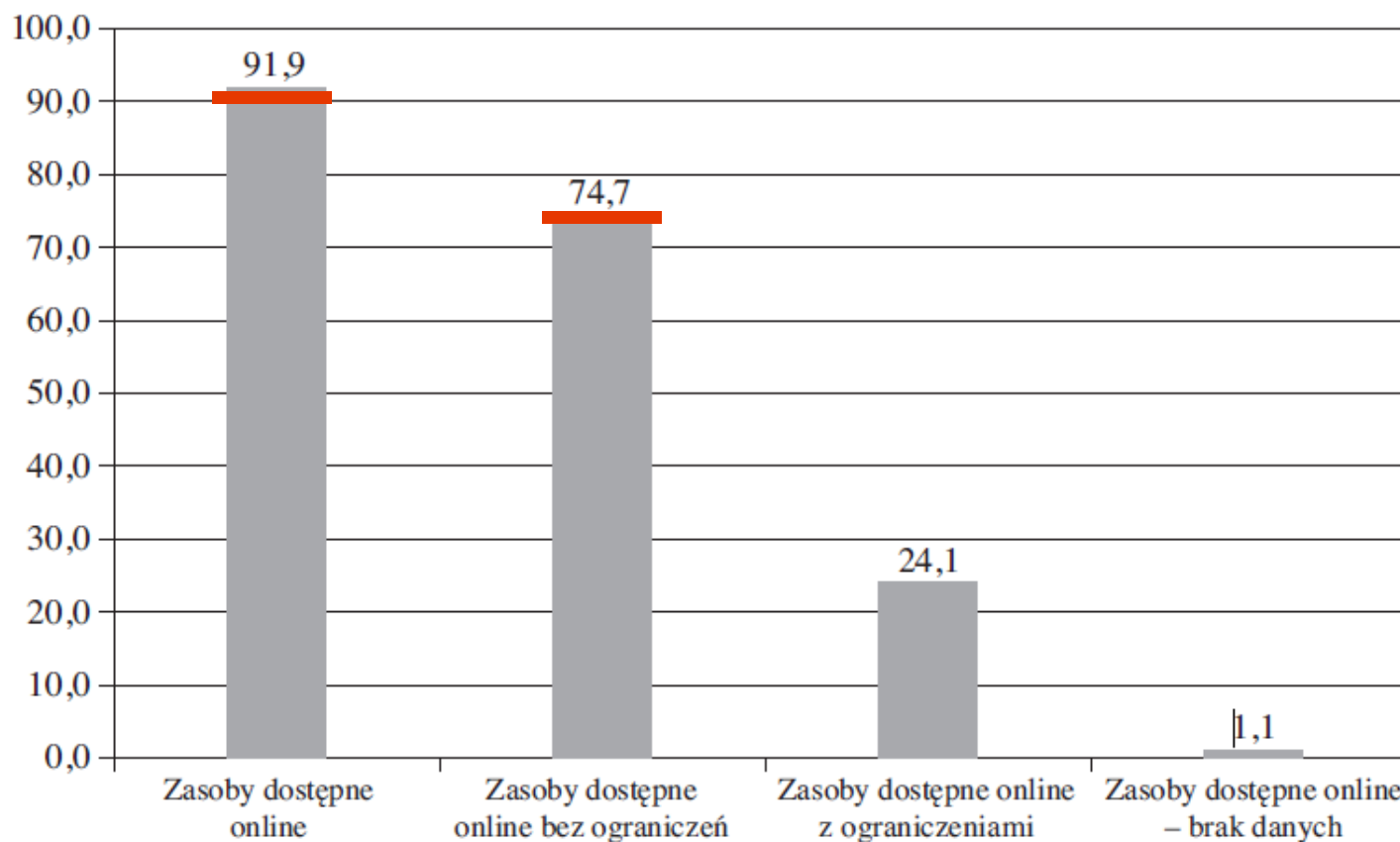




# Sytuacja w Polsce - w świetle badań

## Instytucje

Wykres 1. Odsetek zasobów danych udostępnianych online przez instytucje badawcze



Źródło: Adam Jachimczyk, *Otwarte dane badawcze. Casus polskich instytucji badawczych*, *Zagadnienia Naukoznawstwa* 4 (206), 2015, s. 413







# Sytuacja w Polsce - w świetle badań

## Instytucje

Gdzie są przechowywane dane uzyskane w wyniku prac badawczych w Państwa instytucji?				
N = 164 (79,23%)		BO = 43 (20,77%)	Suma = 207 (100%)	Liczba wskazań = 262
Odpowiedź	zbiorcze	szkoły wyższe	instytuty badawcze	instytuty PAN
na stanowiskach pracy poszczególnych pracowników	129 (78,66%)	28 (77,78%)	52 (83,87%)	26 (89,66%)
na centralnym serwerze instytucji	71 (43,29%)	9 (25,00%)	34 (54,84%)	12 (41,38%)
w otwartym centralnym repozytorium cyfrowym instytucji	19 (11,59%)	2 (5,56%)	7 (11,29%)	4 (13,79%)
w zewnętrznym otwartym repozytorium	15 (9,15%)	5 (13,89%)	5 (8,06%)	2 (6,90%)
inne	28 (17,07%)	7 (19,44%)	12 (19,35%)	5 (17,24%)
liczba respondentów	164	36	62	29

Źródło: Edyta Strzelczyk, *Enhancing scholarly communication: National initiatives to manage research data in the V4 countries. Raport z badań polskich instytucji naukowych i badawczo-rozwojowych (materiał niepublikowany)*





# Korzyści z otwierania danych



niepowielanie  
pracy



oszczędność  
kosztów



misja publiczna



wzrost wartości  
marki



nowe szanse  
finansowania



wzbogacenie danych



nowy materiał  
badawczy



wyszukiwalność



większa liczba  
kanałów dostępu  
dla użytkowników



zwiększenie  
trafności



samokorygowanie  
nauki



wykorzystanie  
przez innych



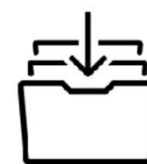
współpraca



rozwój  
komunikacji



nowi odbiorcy



bezpieczna  
archiwizacja



innowacje





# Otwarta nauka

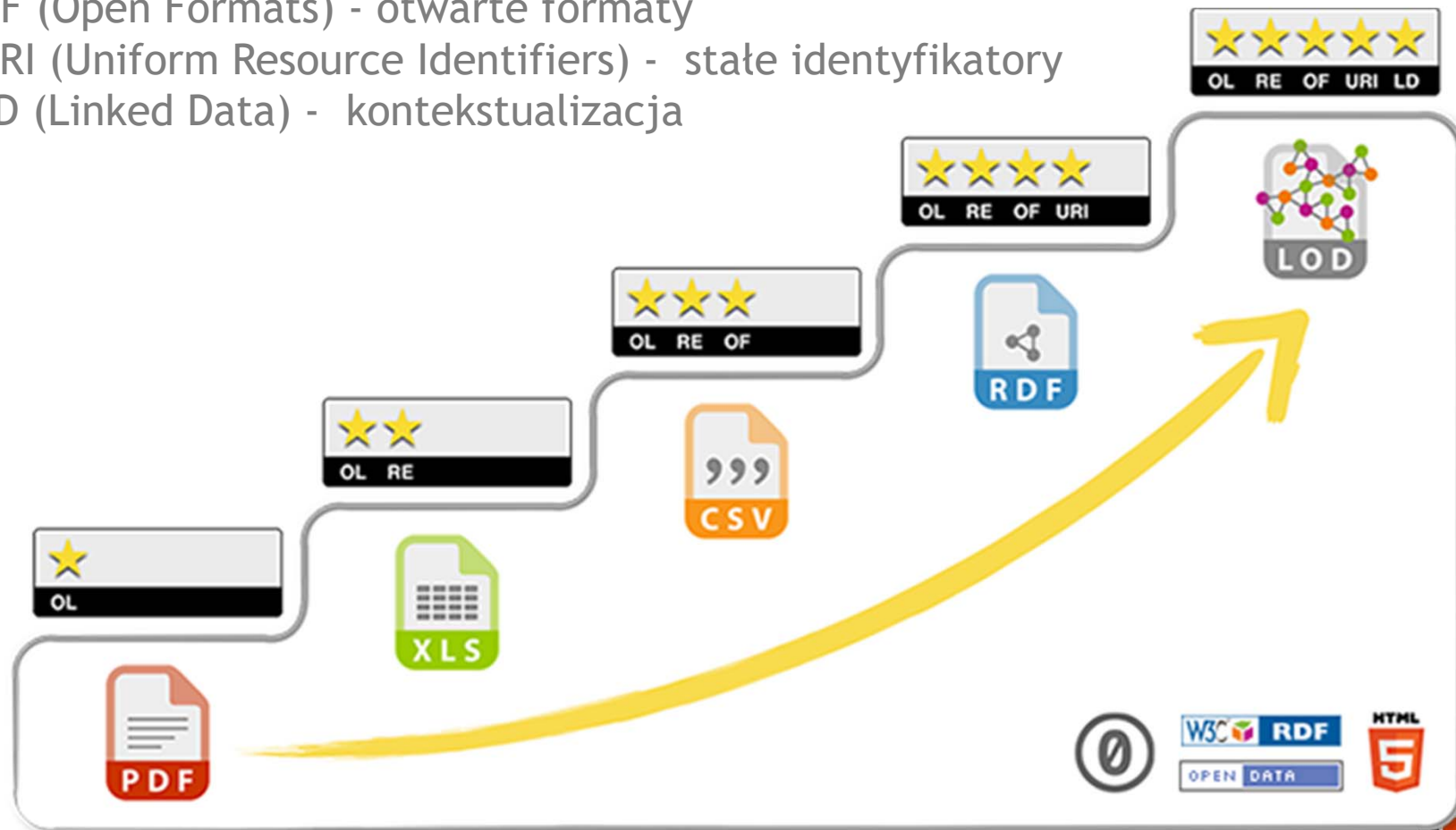
OL (Open Licenses) - otwarte licencje

RE (Reusable) - gotowe do ponownego przetwarzania

OF (Open Formats) - otwarte formaty

URI (Uniform Resource Identifiers) - stałe identyfikatory

LD (Linked Data) - kontekstualizacja

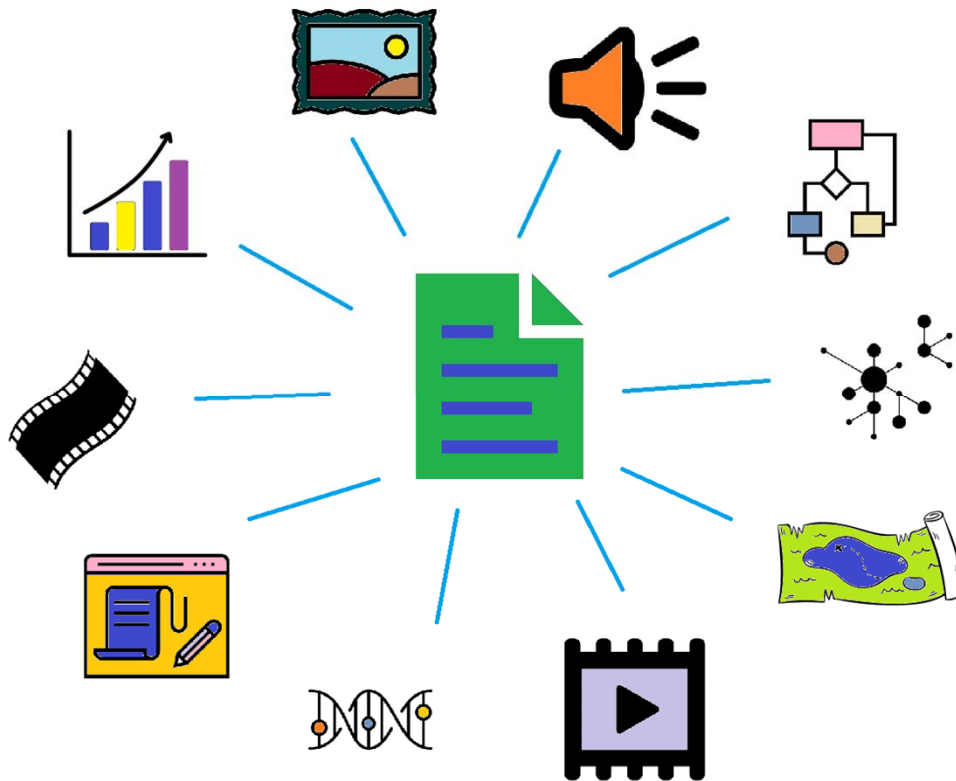


Źródło: <http://5stardata.info/en/>



# Enhanced publications

## Wzmocnione publikacje

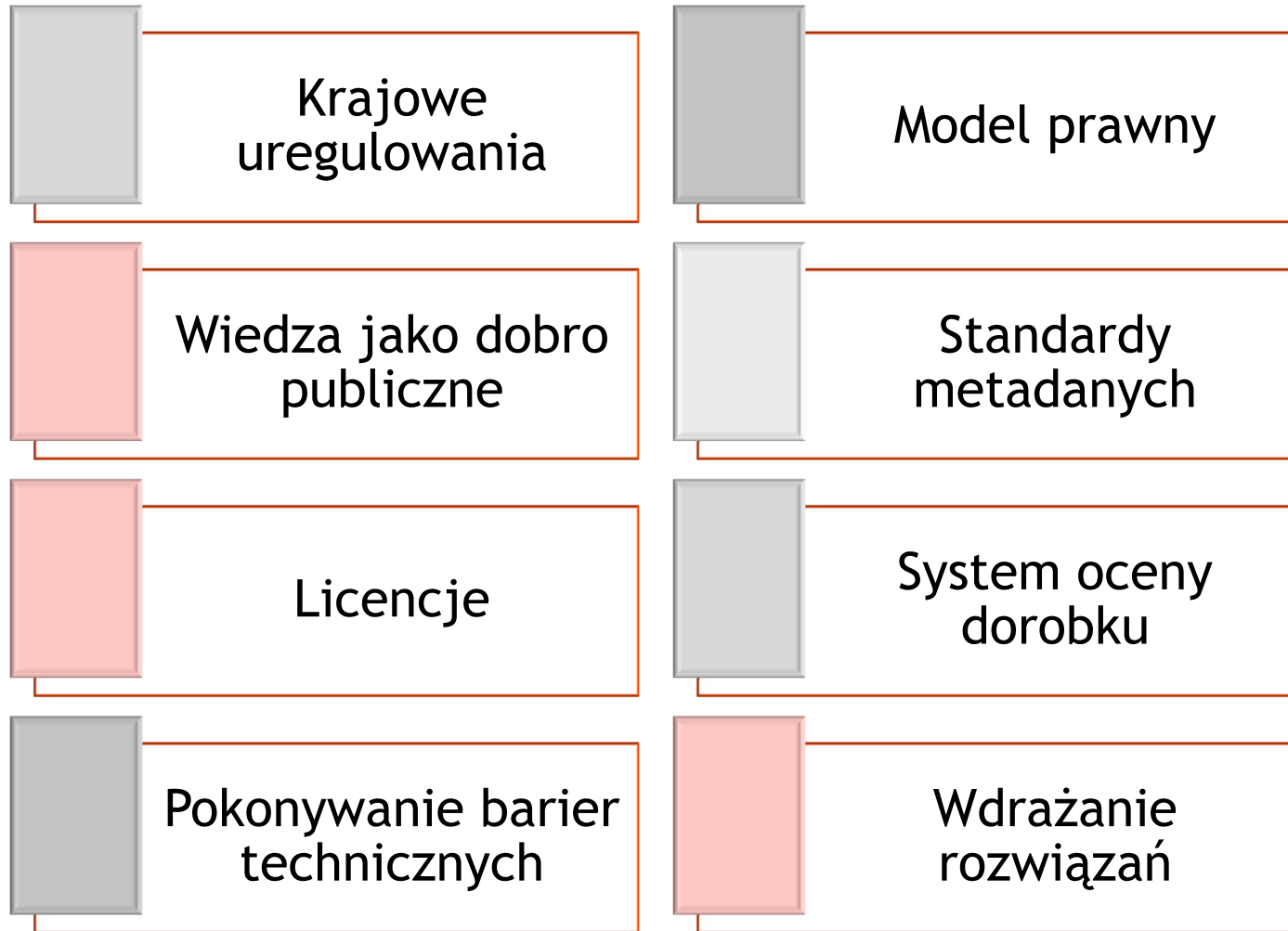


- dane badawcze (załączone lub wskazane miejsce),
- dodatkowe materiały,
- interaktywne treści,
- linki do cytowanych publikacji,
- publikacje nas cytujące,
- komentarze, dyskusje, recenzje,
- relacje z innymi publikacjami, plikami,
- nawigacja,
- metadane,
- otwarte formaty,
- możliwość rozwoju





# Działania na przyszłość





# Bibliografia

- BEDNAREK-MICHALSKA, B. Repozytoria surowych danych – dlaczego biblioteki powinny je znać? *Biuletyn EBIB* [online]. 2012, nr 8 (135). [Dostęp 22.06.2017]. Dostęp w: <http://open.ebib.pl/ojs/index.php/ebib/article/view/152/292>
- GUIBAULT, L., WIEBEGO, A. *Otwarte, czyli bezpieczne. Analiza prawnych form ochrony danych badawczych i zalecenia w kwestii dostępu i korzystania* [online] Kraków, 2014. [Dostęp 22.06.2017]. Dostęp w: <http://jbc.bj.uj.edu.pl/dlibra/doccontent?id=283369>
- HOFMOKL, J., TARKOWSKI, A., BEDNAREK-MICHALSKA, B. i in. *Przewodnik po otwartej nauce* [online]. [Dostęp 22.06.2017]. Dostęp w: <https://otwartanauka.pl/wprowadzenie/przewodnik-po-otwartej-nauce>
- JACHIMCZYK, A. Otwarte dane badawcze. Casus polskich instytutów badawczych, *Zagadnienia naukoznawstwa* 2015, nr 4 (206), s. 409-424. ISSN 0044 -1619
- JAMES, L. *Defining Open Data* [online]. [Dostęp 22.06.2017]. Dostęp w: <http://blog.okfn.org/2013/10/03/defining-open-data/>
- KARWOWSKI, M. Otwarte dane - przyszłość humanistyki? W: *Tematy modne w humanistyce. Studia interdyscyplinarne* [online]. Toruń, 2015. [Dostęp 22.06.2017]. Dostęp w: [https://www.google.pl/search?q=tematy+modne+w+humanistyce&ie=utf-8&oe=utf-8&client=firefox-b&gfe\\_rd=cr&ei=6eBLWca4JM6v8wex6JX4BQ](https://www.google.pl/search?q=tematy+modne+w+humanistyce&ie=utf-8&oe=utf-8&client=firefox-b&gfe_rd=cr&ei=6eBLWca4JM6v8wex6JX4BQ)
- KĘDZIERSKA, E. Wzmocnione publikacje: nowy model wsparcia komunikacji naukowej. *Zagadnienia Informacji Naukowej* 2015, nr 1, s. 44-58.
- *Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce* [online]. [Dostęp 22.06.2017]. Dostęp w: [http://www.nauka.gov.pl/g2/oryginal/2015\\_10/9f62cc350837b942e51ae23dd1f23df8.pdf](http://www.nauka.gov.pl/g2/oryginal/2015_10/9f62cc350837b942e51ae23dd1f23df8.pdf)
- *List Ministra Nauki i Szkolnictwa Wyższego z dnia 10.02.2017 r. dotyczący: otwartego dostępu do publikacji naukowych* [online]. [Dostęp 22.06.2017]. Dostęp w: [http://www.nauka.gov.pl/g2/oryginal/2017\\_02/6b8c8c9c87a816fff3bd6e186b194b9e.pdf](http://www.nauka.gov.pl/g2/oryginal/2017_02/6b8c8c9c87a816fff3bd6e186b194b9e.pdf)
- *Otwarte dane w świecie wielkich danych. Porozumienie międzynarodowe. Wersja skrócona* [online]. [Dostęp 22.06.2017]. Dostęp w: [http://www.science-international.org/sites/default/files/reports/otwarte-dane-wersja-skrocona\\_short\\_pl.pdf](http://www.science-international.org/sites/default/files/reports/otwarte-dane-wersja-skrocona_short_pl.pdf)
- PAWEŁOSZEK, I. *Wybrane problemy wdrożenia koncepcji otwartych danych w e-administracji* [online]. [Dostęp 22.06.2017]. Dostęp w: [http://rocznikikae.sgh.waw.pl/p/roczniki\\_kae\\_z33\\_26.pdf](http://rocznikikae.sgh.waw.pl/p/roczniki_kae_z33_26.pdf)
- ROSZKOWSKI, M. *Otwarte dane a obieg informacji w nauce* [online]. [Dostęp 22.06.2017]. Dostęp w: [https://www.researchgate.net/publication/269050524\\_Otwarte\\_dane\\_a\\_obieg\\_informacji\\_w\\_nauce](https://www.researchgate.net/publication/269050524_Otwarte_dane_a_obieg_informacji_w_nauce)
- STARCZEWSKI, M., STĘPIŃSKA-USTASIAK, L., *Otwarty dostęp czy otwarta nauka?* [online]. [Dostęp 22.06.2017]. Dostęp w: <https://depot.ceon.pl/bitstream/handle/123456789/6317/Otwarty%20dost%C4%99p%20czy%20otwarta%20nauka.pdf?sequence=2&isAllowed=y>
- SZCZĘSNY, P. *Otwarta nauka czyli dobre praktyki uczonych* [online]. Toruń: Stowarzyszenie EBIB, 2013. [Dostęp 22.06.2017]. Dostęp w: [http://repozytorium.uwb.edu.pl/jspui/download/Szczesny\\_Otwarta\\_nauka.pdf](http://repozytorium.uwb.edu.pl/jspui/download/Szczesny_Otwarta_nauka.pdf)
- SZPROTA, J. (red.). *Otwarta nauka w Polsce 2014. Diagnoza* [online]. [Dostęp 22.06.2017]. Dostęp w: <http://pon.edu.pl/index.php/nasze-publicacje?pubid=13>
- UHLIR, P. F., SCHRÖDER, P. (2007) Open data for global science. *Data Science Journal* [online] 2007, 17 [Dostęp 22.06.2017]. Dostęp w: <http://datascience.codata.org/articles/abstract/10.2481/dsj.6.OD36/>
- *What is research data?* ANDS Guide [online]. [Dostęp 22.06.2017]. Dostęp w: [http://www.ands.org.au/\\_data/assets/pdf\\_file/0006/731823/What-is-research-data.pdf](http://www.ands.org.au/_data/assets/pdf_file/0006/731823/What-is-research-data.pdf)



Dziękuję za uwagę

