



***Otwarte dane surowe***

# Co to są otwarte dane surowe

Dane surowe to są te dane, które powstają w wyniku badań naukowych i są gromadzone przez naukowców za pomocą różnych metod, standardów, procedur. Każda dziedzina wiedzy zbiera inne dane potrzebne do realizowania badań, mogą to być :

- 1) Dane statystyczne reprezentowane w formie wykresów, tabel, schematów, map.
- 2) Dane socjologiczne reprezentowane w formie ankiet, testów, wywiadów pisemnych czy audio/video, kwestionariuszy, obserwacji.
- 3) Dane geograficzne reprezentowane w formie tabel, map, wykresów, fotografii.
- 4) Dane osobowe reprezentowane w postaci tabel, wykresów, ankiet, fotografii.
- 5) Dane astronomiczne reprezentowane w postaci map, fotografii, zapisu pomiarów, obserwacji, testów.



# Panton Principles

W roku 2009 naukowcy Peter Murray-Rust, Cameron Neylon, Rufus Pollock i John Wilbanks spisali w Cambridge (Wielka Brytania) kilka zasad odnoszących się do prawnej otwartości danych badawczych. Zasady te zostały potem dopracowane przez członków grupy roboczej Open Knowledge Foundation Working Group on Open Data in Science i oficjalnie zaanonsowane w lutym 2010 r. Są one dziś znane pod nazwą *Panton Principles* (<http://pantonprinciples.org/>).



# Panton Principles

Formalnie zaleca się, by wszystkie dane i publikacje wytworzone za publiczne pieniądze od razu przechodziły do domeny publicznej, były dobrem wspólnym. Można wykorzystać do tego licencje Public Domain Dedication lub licencje Creative Commons „Zero”. Jeśli nie uda się wypełnić takiego warunku, to wylicza się inne możliwości, które powinny być spełnione:

Udostępnionym danym powinno towarzyszyć klarowne oświadczenie dotyczące woli i oczekiwań co do ponownego użycia zarówno pojedynczych rekordów, części, jak i całości danych. Takie oświadczenie powinno być precyzyjne, nieodwołalne i oparte na wybranej formule licencji lub całkowitym zwolnieniu z ochrony. Kiedy udostępnia się dane, powinno się formalnie określić zasady ich wykorzystania.

Wiele powszechnie uznanych licencji nie jest odpowiednich dla publikowania otwartych, surowych danych lub zbiorów danych. Licencje, które są stworzone i dostosowane do stosowania w odniesieniu do zasobów sieci, takie jak: Creative Commons, GFDL, GPL, BSD itp., nie są odpowiednie dla danych, a ich użycie jest odradzane. Zaleca się używanie licencji odpowiednich dla danych, które nazywają się Conformant Data Licenses

<http://opendefinition.org/licenses/#Data>.

Nie powinno się stosować licencji, które ograniczają ponowne handlowe wykorzystanie lub ograniczają produkcję dzieł pochodnych, wyłączających stosowanie ich do określonych celów, konkretnych osób czy organizacji — to jest zdecydowanie odradzane. Licencje takie uniemożliwiają skuteczne ich zintegrowanie i ponowne zastosowanie w działalności gospodarczej, gdzie mogłyby zostać wykorzystane z pożytkiem.



OPEN DATA

# Jakie są otwarte dane surowe ?

**Dane mogą być kolekcjonowane tradycyjnymi metodami lub elektronicznie, dane mogą być ilościowe lub jakościowe. Mogą być reprezentowane jako dane liczbowe, teksty, fotografie, pliki audio/video, itp. Źródłem danych może być człowiek lub zwierzę, notatka terenowa, czasopismo, próbka laboratoryjna, obserwacja, itp.**

**Różne dyscypliny posługują się odmiennymi metodami pracy, pojęciami, procedurami, które stanowią o danej dyscyplinie, dlatego zbierane dane bardzo się różnią i ich opisywanie czy formaty są zależne od dyscypliny.**



# Jak gromadzi się surowe dane ?

Dane gromadzone są przez naukowców. Zarządzający danymi musi uwzględnić następujące warunki:

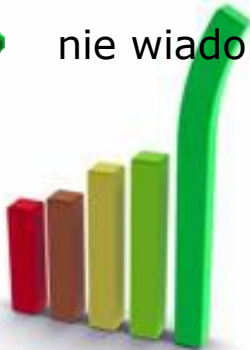
1. Spełnić wymogi ciał finansujących badania i gromadzenie danych.
2. Zapewnić integralność badań i replikację danych.
3. Zapewnić, by dane badawcze i ich opisy były dokładne, kompletne, autentyczne i wiarygodne.
4. Zapewnić, by dane zwiększały efektywność badawczą, zapewnić im odpowiedni zakres upowszechnienia.
5. Gromadzenie danych ma zapewnić oszczędność czasu i środków w długim okresie.
6. Gromadzący dane ma zwiększyć bezpieczeństwo danych i zminimalizować ryzyko ich utraty.
7. Gromadzenie danych ma zapobiec powielaniu wysiłków, umożliwić innym ich wykorzystanie.
8. Gromadzący dane powinien przestrzegać praktyk prowadzonych w przemyśle i handlu.



# Czy warto gromadzić dane?

Sami naukowcy mają wiele powodów, by nie dzielić się danymi:

- dzielenie się danymi oznacza, że wypuszczamy je z rąk, a przecież zbieranie danych zabiera czas i pieniądze;
- tracimy kontrolę nad danymi;
- dane mogą być źle zinterpretowane;
- dane zawierają błędy;
- dane mogą być złe i ktoś to odkryje;
- ktoś mógłby z moich danych wyciągnąć lepsze wnioski, niż ja;
- ktoś szybciej może je wykorzystać do badań czy publikacji niż ja;
- dane wymagają przygotowania opisu i przeformatowania do repozytorium, to zajmuje czas;
- nie wiadomo jak i gdzie składować dane.



# Listy repozytoriów

**DataBib** – katalog repozytoriów surowych danych: <http://databib.org/>  
**DataCite** list of repositories - serwis stworzony dla nadawania numerów DOI danym i kontrolowania cytowań <http://www.datacite.org/repolist/>  
**Global Change Master Directory**  
<http://gcmd.nasa.gov/KeywordSearch/Home.do> - serwis informacyjny tworzony przez Earth Sciences Directorate w NASA.  
**Open Access Directory Data Repositories**  
[http://oad.simmons.edu/oadwiki/Data\\_repositories/](http://oad.simmons.edu/oadwiki/Data_repositories/) - lista tworzona od 2008 r. przez studentów i pracowników Simmons College.  
**Public Data Sets** on Amazon Web Services <http://aws.amazon.com/datasets/> - niekomercyjny serwis Amazona, związany z informowaniem o naukowych bazach danych surowych.





# Repozytoria surowych danych - przykłady



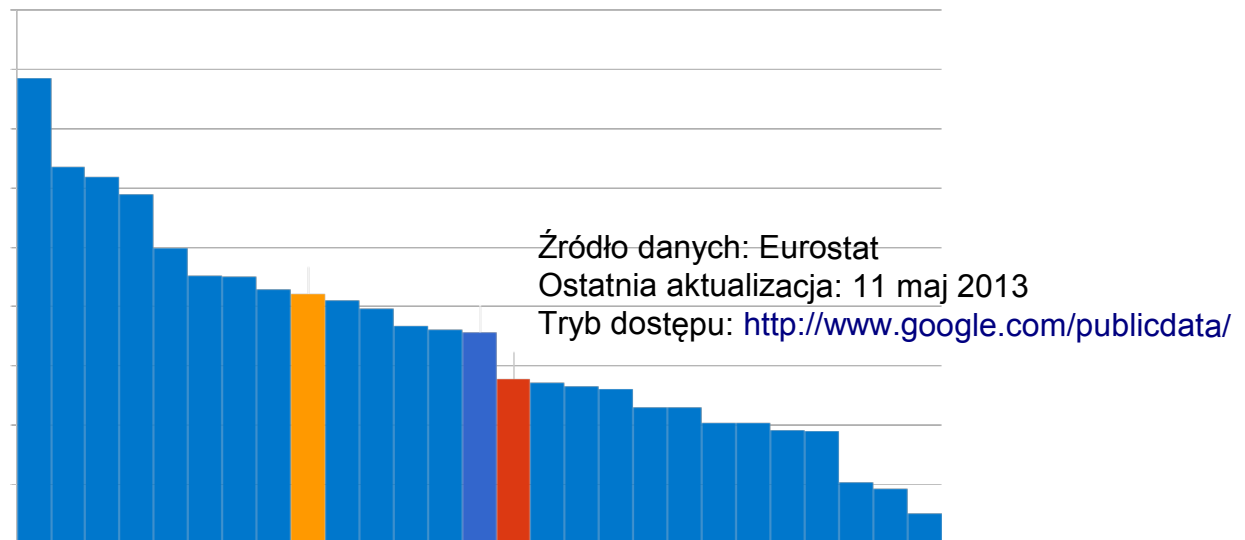
# Jak naukowiec może skorzystać z surowych danych?

- **Przetwarzać dane i wykorzystać do swoich badań (oszczędzając czas na ich osobne gromadzenie).**
- **Zacytować dane w swojej pracy badawczej, dla ilustracji konkretnej tezy (wzbogacić publikację o wiarygodne dane).**
- **Wykorzystać dane do zajęć dydaktycznych ze studentami (wskazać kierunki poszukiwań badawczych).**



# Jak zwykły obywatel może skorzystać z upowszechnienia surowych danych?

**Google Public Data – firma Google na podstawie upowszechnionych za darmo danych Banku Światowego, Eurostatu i innych firm stworzyła eksplorator, który umożliwia ludziom przeglądanie i porównywanie danych statystycznych dotyczących np. zadłużenia państw w Europie:**



# Jaka jest rola biblioteki w kontekście gromadzenia surowych danych?

**Biblioteki mogą pełnić różne role w tym kontekście :**

**1. Mogą doradzać naukowcom jakie i gdzie są repozytoria surowych danych po to, by mogli je składować we właściwym miejscu.**

**2. Mogą wspólnie z konkretnymi instytutami tworzyć repozytorium surowych danych, by wspierać naukowców w ich składowaniu, zabezpieczaniu i upowszechnianiu. Ale należy wziąć pod uwagę fakt, że jest to bardzo trudne zadanie, wymagające wiedzy specjalistycznej.**

**3. Mogą informować o standardach związanych z gromadzeniem konkretnych danych.**





***Dziękuję za uwagę !***

***Bożena Bednarek-Michalska  
Biblioteka Uniwersytecka w Toruniu***