

Otwarte dane w świecie wielkich danych

Porozumienie międzynarodowe
WERSJA SKRÓCONA

Niniejsze porozumienie powstało jako rezultat spotkania „Science International 2015”, pierwszego z serii corocznych spotkań, w których uczestniczą cztery najważniejsze międzynarodowe organizacje naukowe, reprezentujące światową społeczność akademicką na arenie międzynarodowej polityki naukowej (International Council for Science – ICSU, InterAcademy Partnership – IAP, The World Academy of Sciences – TWAS, oraz International Social Science Council – ISSC).

Jako kluczową dziś kwestię globalnej polityki naukowej porozumienie wskazuje szanse i wyzwania związane z rewolucją danych. Zawiera propozycje zasad podstawowych, które powinny zostać przyjęte w odpowiedzi na te szanse i wyzwania. Porozumienie reprezentuje głos społeczności naukowej, odrębny od głosu rządów i organów międzyrządowych, które wskazały otwarte dane jako podstawowy warunek zachowania rygoru badań naukowych i maksymalizacji korzyści dla społeczeństwa wynikających z rewolucji danych, zarówno w krajach rozwiniętych, jak i rozwijających się.

Partnerzy „Science International” będą wspierać dyskusję nad proponowanymi zasadami oraz zachęcać swoich członków, a także inne organy przedstawicielskie nauki działające zarówno na poziomach krajowych jak i na poziomie międzynarodowym, do ich przyjmowania i propagowania.

Dostępna jest również rozszerzona wersja tego porozumienia, która obejmuje głębsze uzasadnienie zawartego w niniejszym dokumencie stanowiska oraz bardziej szczegółowo wypowiada się o praktykach mających na celu wspieranie rozwoju systemów otwartych danych

1. Świat wielkich danych

Rewolucja cyfrowa ostatnich dekad to wydarzenie historyczne na skalę świata, głębsze i bardziej wszechogarniające niż upowszechnienie prasy drukarskiej. Nastąpił bezprecedensowy rozwój w dziedzinie generowania, magazynowania, przetwarzania i natychmiastowego przesyłania danych o ogromnych rozmiarach i ogromnej złożoności, co ma poważne następstwa dla nauki.¹ Szybkość zmian jest zawrotna. W 2003 roku naukowcy ogłosili zakończenie mapowania ludzkiego genomu. Zajęło to ponad 10 lat i kosztowało miliard dolarów - dziś byłoby kwestią dni i pochłonęłoby niewielki ułamek ówczesnych kosztów (tysiąc dolarów). Siłą napędową rewolucji cyfrowej są „wielkie dane” (*Big Data*), gdzie bezprecedensowe ilości danych przepływają przez systemy komputerowe, oraz „szerokie dane” (*Broad Data*), gdzie liczne zbiory danych można łączyć ze sobą na gruncie semantycznym, aby odkrywać ich głębsze znaczenie. „Wielkie” i „szerokie” dane oferują nowe możliwości dla nauk przyrodniczych, społecznych i humanistycznych.

2. Szanse

Ta obfitość danych otwiera dla nauki nowe możliwości: pozwala odkrywać prawidłowości, które dotychczas znajdowały się poza zasięgiem; pozwala łączyć i korelować rozmaite aspekty systemów, aby lepiej zrozumieć ich zachowanie; pozwala opisywać złożoność zjawisk; pozwala także przechodzić od opisu złożonego systemu w określonym stanie do symulacji, które przewidują jego zachowanie dynamiczne. Istnieje wiele obszarów badań, w których takie możliwości są niezwykle ważne: dotyczy to prognozowania pogody i klimatu; badań nad funkcjonowaniem mózgu; badań nad funkcjonowaniem globalnej gospodarki; oceny wydajności produkcji rolnej; prognozowania demograficznego; rozwikływania historii; oraz wielu współczesnych wyzwań o charakterze globalnym, takich jak zmiany w środowisku naturalnym, choroby zakaźne czy masowe migracje, które wymagają wielostronnych analiz i łączenia danych z wielu dyscyplin.

3. Wyzwania

Wykorzystanie tych szans stwarza poważne wyzwania dla sposobu, w jaki prowadzone i organizowane są badania naukowe. Otwarte dane to warunek wstępny niezbędnych zmian.

Imperatyw Otwartych Danych

Podstawową rolą badań finansowanych ze środków publicznych jest poszerzanie zakresu naszej wiedzy i naszego rozumienia, mających kluczowe znaczenie dla wyrażanych sądów, dla innowacji oraz dla dobrobytu społecznego i osobistego. Technologie i procesy rewolucji cyfrowej pozwalają, by dane i koncepcje krążyły w sposób otwarty, szybko i płynnie, za pośrednictwem sieci, pomiędzy wieloma umysłami. Stanowią dzięki temu potężne medium, wspomagające wzrost produktywności i kreatywności w nauce. Aby ta rewolucja społeczna w nauce mogła postępować, musimy przyjąć jako zasadę, że dane zebrane w ramach projektów finansowanych ze środków publicznych powinny po zakończeniu projektu być dostępne publicznie i przeznaczone do ponownego użytku.

Zachowanie zdolności samokorygujących nauki

Otwartość materiału dowodowego (danych), na którym oparte są sformułowane twierdzenia naukowe, to podstawa postępu w nauce. Dzięki dostępowi do danych można poddać kontroli logikę argumentacji oraz powtarzalność obserwacji i eksperymentów, co pozwala potwierdzić lub obalić analizowane twierdzenia. Jest rzeczą konieczną, aby dla każdego publikowanego artykułu zawierającego twierdzenia naukowe dostępne były wykorzystane dane dowodowe, a także powiązane z nimi meta-dane oraz kod programów stosowanych do ich komputerowego przetwarzania. Dzięki temu uda się zachować kluczowy proces samokorygowania nauki. Wykazane w ostatnich czasach w kilku dziedzinach nauki wysokie odsetki opublikowanych wyników naukowych, które nie są powtarzalne, świadczą o silnej potrzebie ożywienia procesów otwierania danych w obecnym świecie wielkich danych. Jednak samo otwarcie danych nie wystarczy. Dane należy otwierać w sposób inteligentny, co oznacza, że muszą być: łatwo wyszukiwalne, dostępne, zrozumiałe, możliwe do oceny i zdadne do ponownego wykorzystania.

Adaptacja rozumowania naukowego

Wiele złożonych relacji, które dziś staramy się uchwycić poprzez powiązane ze sobą dane „wielkie” i „szerokie”, leży daleko poza możliwościami analitycznymi licznych klasycznych metod statystycznych. Aby mieć pewność, że wnioski wyciągnięte z wielkich i szerokich danych są prawidłowe, trzeba sięgać do innych metod matematycznych, w tym do metod topologicznych. Korzystając z dużych ilości danych analiza maszynowa i uczenie maszynowe stają się wszechobecne i zyskują istotny wpływ na odkrycia naukowe. Zdolności poznawcze człowieka nie zawsze pozwalają uchwycić skomplikowane prawidłowości, które są rozpoznawane przez maszyny. Stwarza to poważne problemy związane z interfejsem człowiek-maszyna i każe nam stawiać pytania o rolę uczonego w 21. wieku.

Ograniczenia etyczne

Zasada otwartych danych niesie ze sobą implikacje etyczne dla naukowców i dla tematyki prowadzonych przez nich badań. Może się okazać, że nie pokrywa się ona z interesami naukowców, którzy generują dane, dlatego trzeba opracować nowe sposoby honorowania ich wkładu pracy i nagradzania jej. Należy ponadto chronić prywatność osób, których dotyczą dane. W systemie otwartego dzielenia się danymi, w którym dane są przekazywane kolejnym osobom, ich twórcy tracą kontrolę nad ich przyszłym użytkowaniem, a tymczasem dowiedziono, że procedury anonimizacyjne nie są w stanie zagwarantować bezpieczeństwa informacji osobowych.

Otwarte globalne uczestnictwo

Wielkie dane i otwarte dane mogą przynieść ogromne korzyści krajom mniej zamożnym, a zwłaszcza krajom najsłabiej rozwiniętym (LDC, least developed countries). Jednak systemy nauki w krajach LDC są zazwyczaj ubogie w zasoby. Jeżeli nie będą one uczestniczyć w badaniach opartych na wielkich i otwartych danych, ich odległość od krajów bardziej rozwiniętych będzie gwałtownie rosła w najbliższych latach. Nie będą one w stanie gromadzić, magazynować i udostępniać danych,

¹ Słowo „nauka” jest tu używane w znaczeniu systematycznej organizacji wiedzy, którą można racjonalnie wytłumaczyć i odpowiedzialnie stosować. Podobnie jak w większości języków poza angielskim, obejmuje ono wszystkie dziedziny, w tym nauki humanistyczne i społeczne, oraz dyscypliny określane po angielsku akronimem STEM (nauka, technologia, inżynieria, medycyna).

nie będą mogły uczestniczyć w globalnym przedsięwzięciu badawczym, nie będą mogły jako pełnoprawni partnerzy brać udziału w globalnych wysiłkach w zakresie zmian klimatycznych, ochrony zdrowia i ochrony zasobów, ani też w pełni korzystać z takich działań, w których globalne rozwiązania mogą być osiągnięte jedynie pod warunkiem globalnego udziału. Tak więc zarówno kraje rozwijające się, jak i rozwinięte, mają oczywisty, bezpośredni interes we wspieraniu pełnej mobilizacji potencjału nauki krajów najsłabiej rozwiniętych, a tym samym we wspomaganiu realizacji Celów Zrównoważonego Rozwoju ONZ.

Wykorzystanie szansy

Koncepcję otwartych danych można będzie skutecznie zrealizować tylko wtedy, gdy podejmowane będą systemowe działania na poziomach indywidualnym, dyscyplinarnym, krajowym i międzynarodowym. Choć bowiem nauka jest przedsięwzięciem międzynarodowym, to jej realizacja odbywa się w ramach odrębnych krajowych systemów odpowiedzialności, organizacji i zarządzania. Wszystkie one muszą zareagować, by szanse zostały wykorzystane. Agencje finansujące badania oraz instytucje naukowe powinny finansować i wdrażać praktyki, które ułatwiają badaczom inteligentne otwieranie danych i wspierają procesy związane z otwieraniem danych.

Coraz więcej społeczności naukowych dostrzega korzyści z dzielenia się danymi – dotyczy to dziedzin tak różnych jak językoznawstwo, bioinformatyka i krystalografia chemiczna – i coraz częściej dzięki współpracy międzynarodowej w zakresie dostępu i korzystania z otwartych danych, korzyści te stają się rzeczywistością.

Obowiązki związane z promowaniem i rozwijaniem systemów i procedur, które zapewnią międzynarodowy dostęp do danych, ich interoperacyjność i trwałość, ciąży również na organizacjach międzynarodowych, takich jak Committee for Data in Science and Technology (CODATA), International Council for Science (ICSU), World Data System (WDS) oraz Research Data Alliance (RDA).

Otwarta nauka i wiedza publiczna

Koncepcja „otwartej nauki” zrodziła się z przekonania o potrzebie intensywniejszego dialogu środowiska naukowego ze społeczeństwem i jego silniejszego zaangażowania w rozwiązywanie wielu bieżących problemów poprzez współpracę w określaniu zagadnień oraz w projektowaniu i prowadzeniu badań i stosowaniu ich wyników. Istnieją oczywiście takie sytuacje, w których uzasadnione jest (rozsądnie wprowadzone) ograniczenie otwartości, takie jak ochrona bezpieczeństwa, prywatności i interesu właścicielskiego. Jednak występują również przeciwnie tendencje, prowadzące do prywatyzacji wiedzy, sprzeczne z etosem badań naukowych i podstawową potrzebą człowieka do swobodnego korzystania z idei. Jeśli nauka nie ma ulec takim naciskom, światowa społeczność naukowa musi otwarcie i jednoznacznie bronić zasad otwartych danych, otwartej informacji i otwartej wiedzy.

4. Zasady Otwartych Danych

Tak wielkie jest znaczenie i skala wyzwań, z jakimi konfrontowana jest działalność naukowa w obliczu rewolucji danych, że *Science International* uważa, że należy promować następującą deklarację zasad otwartych danych.

Zobowiązania

Naukowcy

i. Naukowcy finansowani ze środków publicznych mają obowiązek działać na rzecz dobra publicznego poprzez tworzenie i przekazywanie nowej wiedzy, której nieodłączną częścią są dane. Powinni oni takie dane udostępniać innym jak najprędzej po ich wytworzeniu, i w taki sposób, aby mogły być ponownie wykorzystywane, także do nowych celów.

ii. Dane, na podstawie których publikowane są twierdzenia naukowe, powinny być na bieżąco udostępniane publicznie, w formie inteligentnie otwartej.² Umożliwi to rygorystyczną weryfikację rozumowania prowadzącego od danych do twierdzeń naukowych, a także pozwoli na zbadanie powtarzalności przywołanych eksperymentów i obserwacji. Na ile jest to możliwe, dane powinny być przechowywane w godnych zaufania i dobrze zarządzanych repozytoriach, bez zbędnych barier dostępu.

iii. Instytucje badawcze i uniwersytety

mają obowiązek stworzyć warunki korzystne dla otwartych danych. Obejmuje to zapewnienie szkoleń w dziedzinie zarządzania danymi oraz ich przechowywania i analizy, a także zapewnienie odpowiedniego wsparcia technicznego, w tym usług bibliotecznych związanych z zarządzaniem i przechowywaniem danych. Instytucje, które zatrudniają naukowców, i organy, które finansują badania, powinny opracować systemy zachęt zawodowych i kryteria awansu zawodowego dla osób zaangażowanych w procesy otwartych danych. Konsensus co do takich kryteriów jest potrzebny na poziomach krajowych, a najlepiej również na szczeblu międzynarodowym, aby ułatwić rozwój pożądaných wzorców mobilności naukowców. Zgodnie z obecnym duchem umiędzynarodowienia w nauce, uniwersytety i inne instytucje naukowe w krajach rozwiniętych powinny współpracować ze swoimi odpowiednikami w krajach rozwijających się, aby wzmacniać ich kompetencje związane z danymi.

iv. Wydawcy

mają obowiązek udostępniać recenzentom dane na potrzeby procesu recenzyjnego, żądać od autorów otwartego i inteligentnego udostępniania danych z chwilą publikacji opartych na nich prac, oraz wymagać prawidłowego cytowania tych danych. Wydawcy mają również obowiązek dalszego udostępniania dorobku naukowego, poprzez otwarte udostępnianie metadanych oraz przez zapewnienie otwartego dostępu dla maszynowej analizy tekstu i danych (*TDM, text and data mining*).

v. Agencje finansujące badania

powinny uznać koszty działań na rzecz otwartych danych za nieodłączną część kosztów prowadzenia badań, oraz powinny zapewnić odpowiednie środki finansowe i polityczne, aby możliwa była długoterminowa stabilność infrastruktury i repozytoriów. Ocena wpływu badań, w szczególności jeżeli jest oparta na indeksach cytowań, powinna wyraźnie uwzględniać wkład twórców danych.

vi. Stowarzyszenia zawodowe, towarzystwa naukowe i akademie

powinny opracować wytyczne i polityki otwartych danych oraz promować związane z nimi korzyści w taki sposób, który odzwierciedla poznawcze normy i praktyki ich członków.

vii. Biblioteki, archiwa i repozytoria

są odpowiedzialne za rozwój norm technicznych oraz za świadczenie usług dotyczących danych, tak aby dane były dostępne dla tych, którzy chcą z nich korzystać, oraz aby pozostały dostępne w dłuższej perspektywie.

Granice otwartości

viii. Otwartość danych powinna być domyślnym punktem wyjścia w nauce finansowanej ze środków publicznych. Wyjątki powinny być ograniczone do kwestii prywatności, bezpieczeństwa, obronności i użytku komercyjnego w interesie publicznym. Proponowane wyjątki powinny być uzasadniane na poziomie indywidualnych przypadków, a nie zbiorowych wykluczeń.

Praktyki ułatwiające

ix. Cytowanie i źródła

Gdy badacz wykorzystuje w publikacji naukowej dane wytworzone przez innych, powinien je cytować podając ich wytwórcę, pochodzenie oraz stały identyfikator cyfrowy.

x. Interoperacyjność

Zarówno dane badawcze, jak i metadane, dzięki którym dane można ocenić i ponownie wykorzystać, powinny cechować się jak najwyższym stopniem interoperacyjności.

xi. Nieograniczony ponowny użytek

Jeżeli dane badawcze nie należą do domeny publicznej, powinny zostać przygotowane do ponownego wykorzystania poprzez dołączenie oświadczenia o zrzeczeniu się praw lub o udzieleniu nierestrykcyjnej licencji, która wyraźnie wskazuje, że dane mogą być ponownie wykorzystywane przy zachowaniu co najwyżej wymogu podania informacji o ich wytwórcy.

xii. Powiązania

Aby jak najlepiej wykorzystać semantyczną wartość danych, otwarte dane należy, zawsze gdy to możliwe, łączyć z innymi danymi, zależnie od ich treści i kontekstu.

Dokument ten przygotowała grupa robocza ICSU-IAP-ISSC-TWAS:

- **Geoffrey Boulton**, Uniwersytet w Edynburgu i Prezydent CODATA, przewodniczący Grupy Roboczej
- **Dominique Babini**, Uniwersytet w Buenos Aires i CLACSO (przedstawiciel ISSC)
- **Simon Hodson**, Dyrektor Wykonawczy CODATA (przedstawiciel ISSC)
- **Jianhui Li**, Chińska Akademia Nauk, CNIC (przedstawiciel IAP)
- **Tshilidzi Marwala**, Uniwersytet w Johannesburgu (przedstawiciel TWAS)
- **Maria G. N. Musoke**, Uniwersytet Makerere, Uganda (przedstawiciel IAP)
- **Paul F. Uhlir**, Uczony, US National Academy of Sciences (przedstawiciel IAP); niezależny konsultant, polityka i zarządzanie danymi
- **Sally Wyatt**, Uniwersytet w Maastricht, & eHumanities, KNAW (przedstawiciel ISSC)

Rozszerzona wersja porozumienia dostępna:
<http://www.science-international.org>

Kopie papierowe:

International Council for Science (ICSU),
 5 rue Auguste Vacquerie, 75116 Paris, France.



www.icsu.org
www.interacademies.net
www.worldsocialscience.org
www.twas.org

