

W ostatnich czasach koniecznością staje się upowszechnianie danych badawczych, przynajmniej w zakresie niezbędnym do weryfikacji twierdzeń zawartych w publikacji naukowej. Dzieje się tak między innymi ze względu na wymogi instytucji finansujących badania, takich jak na przykład Narodowe Centrum Nauki, czy też uczelni posiadających instytucjonalne polityki otwartości oraz wydawców czasopism naukowych.

Jak przygotować dane badawcze do publikacji

Czy wszystkie dane badawcze mogą i powinny być udostępniane? Już na etapie planowania badań należy zidentyfikować potencjalne zbiory danych (ang. *datasets*) oraz zastanowić się nad zasadnością i celowością ich udostępniania w przyszłości. Dobrze jest też kierować się zasadą, że nasze dane badawcze powinny być „tak otwarte, jak to możliwe – tak zamknięte, jak to konieczne” (ang. *as open as possible, as closed as necessary*). W dalszej kolejności należy zastanowić się nad spodziewaną ilością wytworzonych danych w trakcie procesu badawczego. Pomoże to w ustaleniu, jakie nakłady będą potrzebne do krótko- i długoterminowego przechowywania plików.

Plan zarządzania danymi badawczymi

Zgromadzony materiał należy zawsze poddać selekcji i analizie. Trzeba wziąć pod uwagę wartość naukową zebranych danych, sprawdzić, czy zawierają one wszystkie parametry konieczne do odtworzenia eksperymentu, rozpoznać, czy identyczne zbiory danych już nie istnieją w otwartym dostępie. Należy też zastanowić się, czy koszty przechowywania danych są adekwatne do ich wartości merytorycznej, szybkości

dezaktualizacji informacji w danej dyscyplinie i oczekiwanego naukowo-badawczego stopnia oddziaływania w kolejnych badaniach. Rezultatem tej analizy jest plan zarządzania danymi badawczymi (ang. *data management plan*). Jest to dokument opisujący sposób generowania lub wykorzystywania danych w ramach prowadzonego projektu. W planie zarządzania danymi umieszcza się informacje, jak i jakie dane będą gromadzone, zarządzane, przechowywane i udostępniane w trakcie badań oraz w jaki sposób zostaną udostępnione i zarchiwizowane po zakończeniu projektu badawczego. Taki plan pomaga określić zespołowi badawczemu tryb efektywnego i zrównoważonego zarządzania danymi.

Ponadto rzetelne od samego początku projektu zaplanowanie zarządzania danymi zmniejsza ryzyko ich utraty, naruszenia integralności, złamania zasad poufności lub zderzenia się z innymi zagrożeniami, które mogą spowodować, że dane staną się nieczytelne lub nie nadające się do wykorzystania. Plany zarządzania danymi badawczymi mogą mieć różną postać w zależności od grantodawcy, od opisowych i wielostronicowych po krótkie kwestionariusze.

FAIR – Findable, Accessible, Interoperable, Reusable

Proces udostępniania danych badawczych może być złożony, jeśli badaniom naukowym towarzyszą ogromne ilości różnorodnych plików i zestawień. Aby odpowiednio przygotować i upubliczniać dane badawcze stosuje się zasady FAIR – akronim słów ang. *Findable, Accessible, Interoperable, Reusable*. Zgodnie z zaleceniami FAIR, dane powinny być możliwe do odnalezienia przez ludzi i programy komputerowe. Należy umożliwić łatwy dostęp do danych, bez konieczności użycia specjalnego oprogramowania do ich pobrania i odczytu. Można to osiągnąć przez deponowanie danych badawczych w repozytoriach czy zaprezentowanie ich w *data journals*, czasopismach naukowych publikujących recenzowane artykuły opisujące zbiory danych. Należy udostępniać dane interoperacyjne, czyli przygotowane w czytelnej formie, zgodnym z obowiązującymi standardami i powiązane odnośnikami z innymi zbiorami. Dane badawcze powinny być możliwe do ponownego użycia, czyli dokładnie opisane, zaopatrzone w licencję (najlepiej typu *Creative Commons*), informa-

► c.d. na str. 59

► c.d. ze str. 57

Jak przygotować dane...

cję o autorze czy miejscu powstania. Zbiór zaleceń FAIR jest ciągle rozwijany. Nie wszystkie dane mogą być otwarte, natomiast wszystkie powinny być FAIR.

Aspekty prawne

Ważną kwestią jest rozwiązanie aspektów prawnych związanych z publikowaniem danych. Jeśli w toku badań korzysta się z danych już istniejących, trzeba sprawdzić czy można je dalej udostępniać. Konieczne jest też ustalenie, na jakich zasadach mogą być upowszechniane dane wytworzone w ramach projektu, a także zawarcie stosownych umów regulujących zakres ich dalszego wykorzystywania.

Bezwzględnie należy też usunąć dane wrażliwe, stosując anonimizację danych. Odpowiednie opracowanie danych, klarowna dokumentacja, selekcja i analiza pozwalają usystematyzować zasób, co ułatwi przechowywanie materiałów i usprawni dostęp do nich. Zasady kompletowania danych badawczych opublikowano w broszurze „Selekcja i przygotowanie danych badawczych do udostępnienia” przygotowanej przez Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego (ICM) Uniwersytetu Warszawskiego.

■ Małgorzata Wróblewska
■ Małgorzata Rożniakowska-Kłosińska