

Plan zarządzania danymi (Data Management Plan-DMP) w IChO PAN

1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

1.1. W jaki sposób będą pozyskiwane lub wytwarzane nowe dane lub ponownie wykorzystywane dane już istniejące?

Dane zebrane w ramach proponowanego projektu badawczego będą obejmować opracowane procedury syntetyczne i dane analityczne wszystkich uzyskanych związków. Żaden z syntetyzowanych związków nie będzie uważany za dane możliwe do gromadzenia.

Wszystkie szczegóły reakcji przeprowadzonych podczas projektu zostaną zapisane w elektronicznych dziennikach laboratoryjnych. Dane analityczne będą tworzone w formie cyfrowej za pomocą specjalistycznego oprogramowania służącego do obsługi sprzętu analitycznego. Typowy zestaw danych analitycznych dotyczących danego związku będzie obejmował widma ^1H NMR, ^{13}C NMR, IR, pomiar spektroskopii mas oraz analizę elementarną. Dane te będą gromadzone zarówno na serwerach grupy analitycznej, jak i na komputerach zespołu badawczego w dedykowanym folderze z nazwą projektu. Dane analityczne dla każdej próbki zostaną nazwane zgodnie z określonym systemem nazewnictwa – inicjały właściciela, numer próbki, numer zespołu badawczego.

1.2. Jakie dane (tj. rodzaje, formaty, objętości) będą pozyskiwane lub wytwarzane w projekcie?

Dane uzyskane podczas realizacji projektu będą obejmować dwie kategorie:

1. Opisy warunków i parametrów reakcji chemicznych (pliki PDF, 100 MB).
2. Dane spektroskopowe i ogólna charakterystyka wszystkich związków wytwarzanych podczas projektu (surowe dane FID itp., pliki PDF, 5-10 GB).

Wszystkie uzyskane dane będą przechowywane w formie elektronicznej z, jeśli to konieczne, dodatkowymi kopiami papierowymi notatek laboratoryjnych lub danych analitycznych.

2. Dokumentacja i jakość danych.

2.1. Jakie metadane i dokumentacja (np. metodologia oraz sposoby pozyskiwania i organizacji danych) będą towarzyszyć danym w projekcie?

Surowe dane instrumentalne będą zorganizowane w sposób specyficzny dla danego oprogramowania i przechowywane w odpowiednio oznaczonych folderach – z nazwą konkretnego projektu. Ponadto, do zestawu danych będą dołączone informacje o dacie eksperymentu, szczegółach pozyskania konkretnych wyników, nazwach próbek oraz osobach wykonujących analizy. Zeskanowane pliki PDF, notatki lub skany dzienników laboratoryjnych, jeśli to konieczne, będą opatrzone krótkimi notatkami. Ostatecznemu zestawowi danych zdeponowanemu w instytucjonalnym repozytorium danych towarzyszyć będzie plik README zawierający listę zawartości innych plików i przedstawiający używaną konwencję nazewnictwa plików – inicjały właściciela, numer próbki, numer zespołu badawczego.

2.2. Jakie środki kontroli jakości danych zostaną zastosowane?

Poprawność i powtarzalność wszystkich opracowanych procedur i warunków reakcji zostanie zweryfikowana poprzez powtórzenie eksperymentów. Dane analityczne będą uzyskiwane przez

niezależnych, wykwalifikowanych członków zespołu analitycznego przy użyciu wstępnie skalibrowanego sprzętu, zatwierdzonych metod analitycznych i oprogramowania dostarczanego przez zweryfikowanych dostawców. Ręcznie wprowadzone dane zostaną sprawdzone przez innego członka zespołu, aby uniknąć przypadkowych błędów. W związku z tym ryzyko uzyskania nieprawidłowych danych lub wprowadzania błędów podczas ich gromadzenia zostanie zminimalizowane.

3. Przechowywanie i tworzenie kopii zapasowych podczas badań

3.1. W jaki sposób w trakcie projektu będą przechowywane dane i metadane? W jaki sposób będą tworzone ich kopie zapasowe?

Wszystkie dane będą przechowywane w formacie elektronicznym na dyskach twardych należących do naukowców zaangażowanych w projekt i serwerach grupy analitycznej (dotyczy danych analitycznych). Kopie zapasowe będą systematycznie zapisywane (co najmniej raz w miesiącu) na serwerze NAS Instytutu. Zespół badawczy będzie miał dostęp do maksymalnie 1 TB pamięci masowej na serwerze NAS Instytutu, którą można w razie potrzeby rozszerzyć. Zespół IT będzie odpowiedzialny za tworzenie kopii zapasowych i odzyskiwanie danych w przypadku jakichkolwiek uszkodzeń. Papierowe kopie notatek laboratoryjnych lub danych analitycznych zostaną utworzone tylko wtedy, gdy zajdzie taka potrzeba, a każdy dokument zostanie zeskanowany w celu zapewnienia właściwego przechowywania. W przypadku pracy zdalnej wszystkie wygenerowane dane będą wysyłane do serwera NAS Instytutu za pośrednictwem szyfrowanego połączenia z serwerem.

3.2. W jaki sposób zostanie zapewnione bezpieczeństwo i ochrona danych wrażliwych w okresie trwania projektu?

Wszystkie komputery używane do rejestrowania, przechowywania lub analizowania danych mają odpowiednie oprogramowanie zabezpieczające i ochronę antywirusową. Używane komputery oraz serwery kopii zapasowych są chronione hasłem i szyfrowane. Każdy członek zespołu projektowego będzie miał dostęp do danych przy użyciu hasła. Aby zapewnić ochronę danych, wszystkie pliki będą udostępniane za pośrednictwem grup prywatnych w systemie przechowywania i udostępniania w chmurze lub na zewnętrznych dyskach twardych, a nie wysyłane pocztą. W przypadku projektów we współpracy wszystkie istotne dane będą przesyłane do systemu przechowywania i udostępniania w chmurze, a zaangażowane strony będą miały dostęp do nich za pośrednictwem kont chronionych hasłem.

4. Wymogi prawne, kodeksy postępowania.

4.1. W jaki sposób zostanie zapewniona zgodność z przepisami dotyczącymi danych osobowych i bezpieczeństwa danych?

Administratorem danych osobowych członków grupy badawczej jest IChO PAN. Instytut będzie przetwarzać dane, za zgodą członków grupy badawczej, na podstawie ich pisemnych oświadczeń, w celu prowadzenia działalności naukowej i badawczej, uzasadnionego interesu IChO PAN i przepisów prawnych, w zależności od okoliczności. Więcej informacji można znaleźć na stronie internetowej IChO PAN (www.icho.edu.pl/rodo).

4.2. W jaki sposób będą zarządzane inne kwestie prawne, takie jak prawa własności intelektualnej i prawa własności? Jakże przepisy znajdują w tym przypadku zastosowanie?

Obowiązujące przepisy znajdują się w "Regulaminie ochrony i korzystania z własności intelektualnej w Instytucie Chemii Organicznej PAN". Twórca wyników posiada osobiste prawa autorskie, w szczególności prawo do autorstwa wyniku, prawo do nienaruszalności treści i formy wyniku, ich wykorzystania oraz prawo do oznaczenia wyniku nazwiskiem autora.

5. Udostępnianie i długotrwałe przechowywanie danych.

5.1. Kiedy i w jaki sposób będą udostępniane dane z projektu? Czy istnieją ewentualne ograniczenia i zakazy dotyczące ich udostępniania?

Zasadniczo dane uzyskane w ramach projektu nie będą udostępniane społecznościom naukowym do czasu przyjęcia publikacji naukowej lub zatwierdzenia wniosku patentowego. Wyniki zostaną opublikowane w czasopiśmie o wysokim wpływie (IF), a dane analityczne i szczegóły dotyczące warunków reakcji będą dostępne bezpłatnie w formie informacji uzupełniających. Ponadto wyniki naukowe projektu będą rozpowszechniane przez Open Access za pomocą jednej z dwóch możliwych ścieżek:

- Złota ścieżka — publikowanie bezpośrednio w czasopiśmie o otwartym dostępie;
- Zielona ścieżka – złożenie kopii już opublikowanej pracy w otwartym archiwum lub repozytorium.

Otwarty dostęp do publikacji naukowych zostanie zapewniony niezwłocznie po dacie publikacji.

Wszelkie inne dane, w tym dane surowe, o ile nie naruszają interesów własności intelektualnej, zostaną zdeponowane w otwartym archiwum – Cyfrowe Repozytorium Instytutów Naukowych i będą dostępne bezpłatnie dla wszystkich zainteresowanych stron. Nieprzetworzone dane będą przechowywane przez co najmniej 3 lata po ukończeniu projektu.

5.2. Jak będzie wyglądać selekcja danych przeznaczonych do utrwalenia i gdzie będą one długoterminowo przechowywane (np. repozytorium danych lub archiwum)?

Wszelkie możliwe do opublikowania wyniki lub materiały, a także wszystkie dane, które leżą u podstaw zgłoszenia patentowego lub publikacji naukowych, będą długotrwale przechowywane na serwerze NAS Instytutu. Wybór odpowiednich danych zostanie dokonany przez Kierownika Projektu. Dane w repozytorium będą przechowywane zgodnie z zasadami jednostki finansującej i Instytutu Chemii Organicznej PAN.

5.3. Jakie metody lub oprogramowanie umożliwiają dostęp do danych i korzystanie z nich?

Wszystkie dane będą przechowywane w standardowych formatach (doc, pdf, fid, cif itp.) dostępnych przy użyciu standardowego oprogramowania biurowego i programów przeznaczonych do analitycznego przetwarzania danych. Jeśli to możliwe, dane będą eksportowane do formatów dostępnych przez oprogramowanie typu Open Source, np. Spinworks, Chem Sketch, Mercury. Takie oprogramowanie jest stosowane praktycznie w każdej grupie badawczej prowadzącej badania w dziedzinie syntezy organicznej, która jest przedmiotem przedłożonej propozycji badawczej.

5.4. W jaki sposób będzie zapewnione zastosowanie unikatowego i trwałego identyfikatora (takiego jak identyfikator dokumentu elektronicznego (DOI)) do każdego zbioru danych?

Nie przewiduje się przypisywania numerów DOI do poszczególnych zbiorów danych. Numery DOI zostaną przypisane do artykułów lub komunikatów publikowanych w czasopiśmie naukowych w ramach projektu. Procedury i dane analityczne dotyczące konkretnej publikacji będą gromadzone w odpowiednio nazwanych folderach, które po zaakceptowaniu artykułu zostaną dodatkowo opatrzone numerem DOI.

6. Zadania związane z zarządzaniem danymi oraz zasoby

6.1. Kto będzie odpowiadał za zarządzanie danymi (tj. kto będzie ich opiekunem)?

Odpowiedzialność za dane wytwarzane lokalnie podczas realizacji planowanych prac badawczych spoczywa na Kierowniku Projektu (PI). Każda osoba odpowiedzialna za konkretne zadanie będzie również odpowiedzialna za właściwe zarządzanie danymi w zakresie swoich zadań. PI będzie odpowiedzialny za przygotowanie plików końcowych i ich przesłanie do repozytorium instytucjonalnego wraz z metadanymi. Dane przesyłane do Cyfrowego Repozytorium Instytutów Naukowych będą nadzorowane przez Pełnomocnika Dyrektora IChO PAN ds. otwartego dostępu do publikacji naukowych i danych badawczych. Pracownicy IT Instytutu będą zarządzać serwerem NAS, na którym będą przechowywane kopie zapasowe.

W przypadku projektów opartych na współpracy każda ze stron będzie odpowiedzialna za dane wygenerowane w ich instytucji.

6.2. Jakie zasoby (na przykład finansowe i czasowe) zostaną przeznaczone na cele zarządzania danymi i zagwarantowania zasad FAIR?

IChO PAN i Kierownik Projektu są wyposażeni w wymagany sprzęt i oprogramowanie niezbędne do wdrożenia planu zarządzania danymi. Środki na pokrycie depozytu zbioru danych w repozytorium instytucjonalnym zostaną przekazane z kosztów pośrednich przedłożonego projektu badawczego.

Przestrzeganie zasad FAIR będzie nadzorowane przez Pełnomocnika Dyrektora ds. otwartego dostępu do publikacji naukowych i danych badawczych.