

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 zbierane w ramach proponowanego projektu badawczego będą obejmować opracowane procedury syntetyczne oraz dane analityczne wszystkich otrzymanych związków. Będą one rejestrowane w elektronicznych dziennikach laboratoryjnych, a dane analityczne będą generowane cyfrowo przy użyciu specjalistycznego oprogramowania. Typowe zestawy danych obejmują widma ^1H NMR, ^{13}C NMR, IR, chromatogramy, rotację właściwą oraz widma masowe. Dane będą przechowywane na serwerach grupy analitycznej oraz komputerach zespołu badawczego, uporządkowane w folderach specyficznych dla danego projektu. Dane analityczne dla każdej próbki zostaną nazwane zgodnie z określonym systemem nazewnictwa – inicjały właściciela, numer próbki, numer zespołu badawczego.

Istniejące dane, w tym opublikowane zestawy widm, protokoły reakcji oraz wcześniej zgromadzone dane wewnętrzne z pokrewnych badań, zostaną przeanalizowane i zintegrowane. Do porównywania wyników oraz optymalizacji metod analitycznych zostaną wykorzystane publiczne bazy danych, takie jak Reaxys i SciFinder.

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 wygenerowane dane będą przechowywane w formacie elektronicznym, z możliwością wykonania dodatkowych kopii papierowych notatek laboratoryjnych lub zbiorów widm, jeśli zajdzie taka potrzeba.

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. Zestaw danych będzie zawierał metadane, takie jak 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. Dane wprowadzane ręcznie zostaną dodatkowo skontrolowane przez innego członka zespołu w celu uniknięcia błędów. Podjęte środki znacząco zminimalizują ryzyko wystąpienia niepoprawnych danych lub stroniczości podczas ich gromadzenia.

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 znajdzie 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? Jakie 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 wyników, prawo do nienaruszalności treści i formy wyników, ich wykorzystania oraz prawo do oznaczenia wyników 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?

Dane wygenerowane w ramach projektu nie będą udostępniane społeczności naukowej do momentu zaakceptowania publikacji lub zatwierdzenia zgłoszenia patentowego. Wyniki zostaną opublikowane w czasopiśmie o wysokim wpływie (IF), a dane analityczne oraz warunki reakcji zostaną udostępnione bezpłatnie jako materiały uzupełniające.

Wyniki naukowe będą rozpowszechniane w trybie Open Access jedną z dwóch ścieżek:

- Złota ścieżka – bezpośrednia publikacja w czasopiśmie Open Access,
- Zielona ścieżka – zdeponowanie kopii opublikowanej pracy w otwartym archiwum lub repozytorium.

Instytut korzysta z OPEN, Repozytorium Otwartych Publikacji Naukowych Uniwersytetu Warszawskiego. Dostęp do publikacji w trybie Open Access będzie możliwy natychmiast po ich opublikowaniu.

Pozostałe dane, w tym surowe dane, które nie naruszają własności intelektualnej, zostaną zdeponowane w RepOD, otwartym repozytorium danych badawczych prowadzonym przez tę samą instytucję. Surowe dane będą przechowywane przez co najmniej 3 lata po zakończeniu 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 długotrwałe 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 kompatybilnych z oprogramowaniem open-source, np. Spinworks, Chem Sketch, Mercury. Narzędzia te są powszechnie wykorzystywane przez grupy badawcze zajmujące się syntezami organicznymi, które stanowią główny obszar proponowanego projektu badawczego.

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

Numery DOI nie będą przypisywane indywidualnym zestawom danych. Zamiast tego numery DOI będą przypisywane artykułom lub komunikatom opublikowanym w czasopiśmie naukowym w ramach projektu. Procedury i dane analityczne związane z każdą publikacją będą uporządkowane w odpowiednio nazwanych folderach, a po zaakceptowaniu artykułu folderom tym zostanie przypisany numer DOI.

Publikacje zdeponowane w repozytorium **OPEN** będą identyfikowane tym samym numerem DOI, co odpowiadająca im publikacja. Zestawy danych zdeponowane w repozytorium **RepOD** otrzymają unikalny numer DOI.

6. Zadania związane z zarządzaniem danymi oraz zasoby

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

Odpowiedzialność za zarządzanie danymi generowanymi lokalnie w trakcie realizacji planowanych prac badawczych spoczywa na Kierowniku Projektu (PI). Każda osoba odpowiedzialna za konkretną część zadania będzie również odpowiedzialna za zarządzanie danymi w ramach tego zadania. Kierownik Projektu będzie nadzorować przygotowanie ostatecznych plików oraz ich przesłanie do repozytorium instytucjonalnego, wraz z niezbędnymi metadanymi. Przesyłanie danych do otwartych archiwów i repozytoriów będzie 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 oraz ostateczne dane.

W przypadku projektów współpracy, każda instytucja partnerska będzie odpowiedzialna za zarządzanie danymi generowanymi w ramach swojej organizacji.

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

Instytut Chemii Organicznej PAN oraz Kierownik Projektu (PI) dysponują niezbędnymi zasobami sprzętowymi i programowymi, aby skutecznie wdrożyć plan zarządzania danymi. Środki finansowe na ostateczne zdeponowanie zestawu danych w repozytorium instytucjonalnym zostaną przydzielone z kosztów pośrednich złożonego projektu badawczego.

Jeśli chodzi o czas, Kierownik Projektu oraz odpowiedni członkowie zespołu projektowego poświęcą niezbędny czas, aby zapewnić, że dane będą zarządzane zgodnie z zasadami FAIR. Nadzór nad zgodnością z tymi zasadami będzie sprawować Pełnomocnik Dyrektora IChO PAN ds. otwartego dostępu do publikacji naukowych i danych badawczych, który zapewni, że dane będą odpowiednio przechowywane, dostępne oraz zgodne z normami dotyczącymi łatwości odnajdywania, przetwarzania i ponownego wykorzystania.