



Przechowywanie i archiwizacja danych badawczych

1. Cel instrukcji

Celem niniejszej instrukcji jest przedstawienie zasad postępowania z danymi badawczymi gromadzonymi lub wytwarzanymi w ramach badań naukowych prowadzonych w Instytucie Rozrodu Zwierząt i Badań Żywności Polskiej Akademii Nauk.

2. Definicje i skróty

Skróty/Używane zwroty	Definicje
IRZiBŻ	Instytut Rozrodu Zwierząt i Badań Żywności
PAN	Polska Akademia Nauk
RDM	Research Data Management (Zarządzanie Danymi Badawczymi)
Serwer RDM	serwer do przechowywania danych
RDM Hot	część serwera dedykowana do przechowywania danych badawczych surowych i przetworzonych i innych danych, które podlegają bieżącemu przetwarzaniu
RDM Cold	część serwera dedykowana do przechowywania danych badawczych surowych i przetworzonych, które nie podlegają dalszemu przetwarzaniu
Hot data	dane surowe i dane przetworzone (dane bieżące, niezarchiwizowane).
Cold data	dane archiwalne, do których dostęp nie jest natychmiastowy (repozytorium lub archiwum/serwer/chmura/dysk zewnętrzny) i są gwarancją autentyczności danych. Dostęp do tych danych jest ograniczony.
Lider projektu (LP)	osoba bezpośrednio odpowiedzialna za badawczy projekt naukowy realizowany w IRZiBŻ PAN. Jest to osoba odpowiedzialna za prowadzenie badań, m.in. osiągnięcie założonych celów projektu zgodnie z zapisami w umowie z instytucją finansującą badania (np.: NCN, NCBiR) lub zlecającym badania.
Data Steward (DS)	osoba, która wspiera badaczy we wszelkich kwestiach związanych z zarządzaniem danymi i ich udostępnianiem. Data Steward jest również odpowiedzialny za realizację polityk wykorzystania danych i bezpieczeństwa określonych w ramach instytucjonalnej inicjatywy zarządzania danymi.
Zakończenie projektu	data zakończenia projektu lub faktyczne zaprzestanie prac nad projektem i/lub publikacja danych naukowych i/lub zdeponowanie ich w repozytorium.
Metadane	ustrukturyzowane informacje opisujące, tłumaczące, lokalizujące i ułatwiające odnalezienie, wykorzystanie lub zarządzanie zasobem informacji. Metadane są kluczem do uzyskania dostępu do danych badawczych, ich zrozumienia i ponownego wykorzystania.

3. Zakres stosowania

Dokument obowiązuje w IRZiBŻ PAN w Olsztynie.

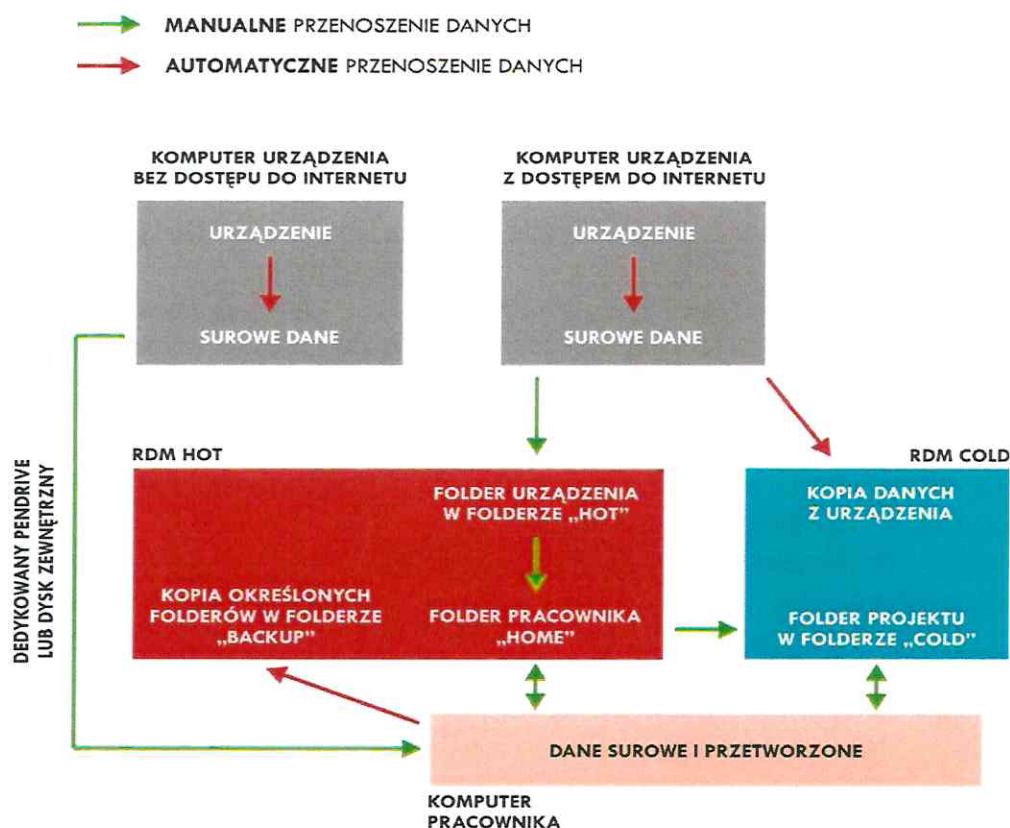
	Imię i nazwisko	Data akceptacji	Podpis
Opracował:	Dr. inż. Michał Janiak	02.02.2023	
Sprawdził:	Beata Marmulewska	02.02.2023	
Zatwierdził:	Prof. dr hab. Mariusz K. Piskula	02.02.2023	

4. Odpowiedzialność

Rola w procesie	Odpowiedzialność
Lider projektu	a. przygotowanie oraz realizacja projektu zgodnie z wytycznymi;
	b. nadzór nad procesem pozyskiwania danych surowych, ich przetworzenia aż do momentu publikacji;
	c. planowanie prac projektowych w taki sposób, aby zapewnić bezpieczeństwo danych w tym określenie sposobu przechowywania i przesyłania danych badawczych, zgodnie z dobrymi praktykami i regulacjami wewnętrznymi IRZiBŻ;
	d. po uzgodnieniu z DS przesłanie administratorowi IT danych na temat ścieżki folderów katalogu projektowego na <i>RDM Cold</i> oraz listy osób mających dostęp do katalogu głównego i/lub poszczególnych folderów;
	e. podejmowanie decyzji, które dane są „ <i>hot data</i> ” a które „ <i>cold data</i> ”;
	f. po zakończeniu projektu kontakt z <i>Data Stewardem</i> w celu przekazania informacji dot. zarchiwowania danych na serwerze Instytutu i udostępnienia danych badawczych w repozytorium lub we wskazanym w projekcie (Planie Zarządzania Danymi) innym miejscu;
	g. zgłaszanie incydentów związanych z bezpieczeństwem danych do administratora IT oraz DS;
	h. dostosowanie mechanizmów kopiowania danych w komputerze do obowiązujących zasad;
Członkowie Zespołu projektowego	a. prowadzenie prac badawczych w sposób zapewniający zgodność z obowiązującymi regulacjami zewnętrznymi i wewnętrznymi;
	b. zgłaszanie incydentów związanych z bezpieczeństwem danych do administratora IT oraz DS;
	c. tworzenie kopii zapasowych danych będących rezultatem działalności Uczestnika/Wykonawcy w projekcie.
Data Steward	a. nadzór nad zapisami i archiwizacją „ <i>cold data</i> ” na serwerze Instytutu;
	b. zgłaszanie do IT próśb dostępu do „ <i>cold data</i> ” dla naukowców;
	c. po zakończeniu czasu archiwizacji dla „ <i>cold data</i> ” (wynoszącym 10 lat) i uzyskaniu zgody Lidera projektu informuje administratora IT o zgodzie na usunięcie danych;
Administrator IT	a. zarządzanie kopiami danych z komputerów;
	b. tworzenie kopii zapasowych danych znajdujących się na serwerze <i>RDM Cold</i> ;
	c. tworzenie lokalizacji danych dla poszczególnych projektów;
	d. nadawanie uprawnień użytkownikom do katalogów i folderów na serwerze RDM;
	e. nadawanie uprawnień dostępu do folderu „ <i>cold data</i> ” dla Lidera projektu po zaakceptowaniu prośby przez <i>Data Stewarda</i> ;
	f. wsparcie techniczne w trakcie trwania projektu.
Osoba odpowiedzialna za stan techniczny urządzenia	a. jest wyznaczana przez Kierownika Zakładu / Laboratorium / etc.;
	b. jest odpowiedzialna za kontakt z Administratorem IT w sprawach związanych z systemem RDM i dotyczących podłączenia urządzenia do RDM;
	c. usuwanie danych badawczych po dwóch miesiącach od ich wytworzenia;

5. Schemat procesu

SCHEMAT ZAPISYWANIA I ARCHIWIZACJI DLA POJEDYNCZEGO PROJEKTU



Rys. 1. Schemat przechowywania i archiwizacji danych badawczych.

6. Opis postępowania

6.1. Informacje ogólne

Instytutowa infrastruktura do przechowywania danych nosi nazwę serwera RDM. Jest on podzielony na część „Hot” i „Cold”. Do tych części dostęp mają naukowcy.

RDM Hot służy do przechowywania danych badawczych pozyskanych w trakcie trwania projektu oraz podlegającym przetwarzaniu i bieżącej archiwizacji (w folderze „home”). Dostęp do tej części serwera jest ciągły i niezależny od czasu trwania projektu. *RDM Cold* (foldery projektowe w folderze „cold”) służy natomiast do przechowywania danych, które podlegają długotrwałej archiwizacji. Umieszczone tam dane źródłowe są gwarancją autentyczności otrzymanych wyników. Po zakończeniu prac nad projektem oraz umieszczeniu w *RDM Cold* ostatecznych, aktualnych wersji przetworzonych danych, dostęp do tej części serwera jest ograniczany w uzgodnieniu z LP. Od tego momentu, dostęp jest przypisany jedynie Data Stewardowi. Pracownicy mogą wnioskować o dostęp zwracając się do DS. Procedura ta ma na celu ochronę danych przed nieautoryzowanym dostępem i modyfikacjami. Gdy czas archiwizacji upłynie (wartość przyjęta to 10 lat), dane po konsultacji z LP są usuwane.

Foldery pracowników na *RDM Hot* (foldery „home”) mają luźniejszą strukturę niż folder projektowy na *RDM Cold*. Niemniej jednak zaleca się stosowanie porządku folderów wymaganego w *RDM Cold* również w przypadku danych umieszczonych w folderach pracowników na *RDM Hot*.

1. IRZiBŻ dba o to, by w trakcie realizacji i po zakończeniu projektu dane badawcze były przechowywane w sposób zapewniający im bezpieczeństwo i uniemożliwiający ich zaginięcie lub dostęp osób niepowołanych.
2. Każdy pracownik ma dostęp do prywatnego folderu na *RDM Hot*, który służy do przechowywania danych. Konto oraz folder tworzy administrator IT.

3. Szczegółowy plan zarządzania danymi badawczymi pozyskiwanymi podczas realizacji projektów, gdzie umowa na finansowanie zadań badawczych wymaga jego powstania na etapie składania projektu określa LP na zasadach finansującej instytucji.

6.2. Pozyskiwanie danych

Dane naukowe mogą być pozyskane w następujący sposób:

- a) Dane pozyskiwane z urzędów laboratoryjnych;
- b) Dane z obserwacji własnych;
- c) Dane pozyskane z innych źródeł, np. wytworzone przez innego naukowca, zdeponowane w repozytorium;
- d) Dane pozyskiwane z usług zewnętrznych, tj. analiza próbek materiału wykonana przez laboratoria zewnętrzne;

W każdym z tych przypadków należy zadbać o to, aby dane nie zostały zmienione w sposób nieautoryzowany. Dotyczy to zarówno danych źródłowych, danych roboczych w trakcie projektu, jak i końcowych.

6.3. Przechowywanie danych podczas trwania projektu (RDM Hot)

1. Wszystkie dane wytwarzane w IRZiBŻ muszą niezwłocznie zostać zabezpieczone poprzez zapisanie ich przynajmniej w dwóch miejscach, w tym w RDM Hot.
2. Dane badawcze zebrane z urzędów mających dostęp do sieci IRZiBŻ muszą zostać niezwłocznie zapisane w przypisanym danemu urządzeniu w folderze „hot”. Następnie, pracownik korzystając z przypisanego mu komputera służbowego przenosi je do własnego folderu na RDM Hot „home”. Po upewnieniu się, że dane badawcze zostały zapisane w folderze „home”, pracownik usuwa dane badawcze z folderu urządzenia w folderze „hot”.
3. W przypadku, gdy urządzenie badawcze nie posiada dostępu do sieci IRZiBŻ, dane należy przenieść używając nośnika zewnętrznego przeznaczonego wyłącznie do tego celu.
4. Surowe dane badawcze zapisane na serwerze RDM Hot należy niezwłocznie zapisać również na serwerze RDM Cold w folderze „cold”, w stosownym subkatalogu projektu/działania naukowego. Sposób korzystania z tej części serwera określa pkt 6.4.
5. Osoba operująca danymi jest zobowiązana do utrzymywania jednoznaczności i schludności w nazewnictwie plików oraz organizacji folderów. Zaleca się poprzedzać część opisową nazwy pliku prefiksem składającym się z daty w formacie „YY.MM.DD_” lub „YYYY.MM.DD_”. Wszelkie oznaczenia dotyczące np. wersji pliku powinny być umieszczane na końcu nazwy pliku.
6. W przypadku, gdy dane badawcze posiadają wersję papierową LP lub osoba przez niego wskazana ma obowiązek porządkowania dokumentacji papierowej w segregatory z właściwym nazewnictwem (lub prowadzenia księgi laboratoryjnej).
7. Dane wyjściowe z aparatury mogą być usuwane z komputera pomiarowego przez osobę odpowiedzialną za stan techniczny danego urządzenia wyznaczaną przez Kierownika Zakładu/Laboratorium po dwóch miesiącach od ich wytworzenia.
8. Pracownik określa folder na przypisanym mu komputerze służbowym, który będzie podlegał automatycznemu tworzeniu kopii zapasowej nad RDM Hot. Kopia taka jest tworzona wyłącznie podczas pracy w siedzibie IRZiBŻ poprzez sieć wewnętrzną.

6.4. Archiwizacja długoterminowa (RDM Cold)

1. LP po uzgodnieniu z DS wnioskuje do działu IT o stworzenie folderu przypisanego danemu projektowi/działaniu naukowemu. Nazwa folderu powinna zostać utworzona wg schematu: **WEWNĘTRZNY-NUMER-PROJEKTU**. W tym folderze tworzone są następnie podfoldery przypisane poszczególnym urządzeniom badawczym wykorzystywanym do wytworzenia danych. Ich lista stanowi załącznik do instrukcji (Zał. nr. 1 Rejestr urzędów badawczych_RDM). LP załącza także listę osób pracujących w projekcie, które będą miały dostęp do katalogu projektowego lub wybranych podfolderów.
2. Struktura ta musi być czytelna i jednoznaczna, a jej utrzymanie jest wymagane przez czas trwania projektu.
3. W powstałych folderach umieszcza się *cold data*. Nazewnictwo plików danych surowych musi pozostać w formie generowanej przez urządzenie badawcze. Natomiast nazwa plików danych przetworzonych musi być

jednoznaczna i zrozumiała dla osoby trzeciej, wymagany jest krótki tytuł. Niedopuszczalne jest stosowanie niezrozumiałych skrótów. Nazwy plików powinny składać się ze słów kluczowych Np.

Pełna nazwa analizy: Oznaczenie zawartości kwercetyny w gryce z uprawy ekstensywnej

Prawidłowa nazwa pliku: „20.05.03_kwercetyna w gryce z uprawy ekstensywnej”,

Przykłady nieprawidłowej nazwy pliku: „2020.05.03_oznaczenie zawartości kwercetyny w gryce z uprawy ekstensywnej” lub „kwer-grk.nowe2”.

4. Aby przetworzone dane badawcze zyskały status *cold* niezbędna jest ich publikacja w postaci artykułu naukowego lub/i złożenie w repozytorium zewnętrznym lub/i decyzja LP. Decyzja LP może być jego/jej subiektywną oceną lub wynikać ze specyficznych uwarunkowań określonych umową na finansowanie badań.
5. Inne dane naukowe niewytworzone na urządzeniach laboratoryjnych (notatki, dane z baz danych, zdjęcia, nagrania audio i wideo itp.) powinny w miarę możliwości zostać zarchiwizowane na serwerze *RDM Cold*, zgodnie z przyjętym przez LP nazewnictwem i strukturą folderów.
6. LP na zakończenie projektu informuje DS, że dane zostały zdeponowane na serwerze Instytutu w folderze *RDM Cold* oraz podaje informacje dotyczące:
 - a) autor
 - b) nazwa projektu i folderu projektowego;
 - c) organizacji przyjętego nazewnictwa folderów;
 - d) okresu archiwizacji dokumentów zgodnie z Planem Zarządzania Danymi lub innymi wytycznymi;
 - e) nazwę repozytorium zewnętrznym gdy dane zostały tam udostępnione.
7. DS zapisuje dostarczone informacje przez LP w Zał. nr. 2 „Rejestr zarchiwizowanych danych badawczych”. Jest on przechowywany na komputerze służbowym *Data Stewarda* oraz na serwerze *RDM Hot* w katalogu home.
8. Po zakończeniu trwania projektu i uzyskania statusu *cold* dostęp do katalogu projektowego na *RDM Cold* dla zespołu projektowego zostaje zmieniony na tylko do odczytu na wniosek LP do DS.
9. Dane umieszczane na *RDM Cold* podlegają tworzeniu kopii zapasowej. Kopia ta jest tworzona na bieżąco i zawiera wszystkie dane, które są tam wprowadzane, nawet gdy zostaną one usunięte z *RDM Cold*. Z tego powodu LP odpowiada za umieszczanie tam wyłącznie danych naukowych podlegających archiwizacji i za nadzór nad tym, aby *RDM Cold* wykorzystywany był przez zespół projektowy do archiwizacji danych źródłowych i przetworzonych nie podlegających już dalszemu przetwarzaniu. Nie należy umieszczać tam danych, nad którymi prace nadal trwają.

6.5. Dostęp do zarchiwizowanych danych po zakończeniu badań i publikacji wyników

1. Na prośbę dowolnego pracownika IRZiBŻ administrator IT, udziela dostępu do odczytu danych na poniższych zasadach:
 - 1.1. Naukowiec zgłasza prośbę o dostęp do DS i podaje:
 - a) do jakich danych potrzebuje dostępu (nazwa projektu, data utworzenia, itp.);
 - b) wskazanie powodu dostępu do danych;
 - c) jeżeli dane nie zostały wytworzone przez osobę wnioskującą, taką prośbę należy rozpatrzyć również z autorem danych badawczych lub z LP.
 - 1.2. DS odnotowuje prośbę w Rejestrze zarchiwizowanych danych badawczych, a następnie dane są udostępniane wnioskodawcy przez Administratora IT. Administrator IT przydziela dostęp do danych na czas określony.
 - 1.3. Rozpatrzenie wniosku o dostęp do danych musi nastąpić maksymalnie w ciągu 24 h od jego wpłynięcia lub w przypadku nieobecności DS w ciągu 24h od jego powrotu do pracy.
2. Po upływie czasu archiwizacji, który został podany przy deponowaniu danych, DS kontaktuje się z LP i ustalają wspólnie przebieg usuwania danych badawczych.
3. Administrator IT usuwa *cold data* na wniosek zaakceptowany przez LP i DS.

6.6. Bezpieczeństwo danych badawczych

1. Do pracy z danymi nie wolno korzystać z rozwiązań chmurowych za wyjątkiem sytuacji gdy w skład grupy badawczej wchodzi osoba spoza Instytutu. Z powodu ryzyka utraty danych w razie zgubienia lub kradzieży nie jest dozwolone przechowywanie i przenoszenie danych na nośnikach zewnętrznych. Wyjątek stanowi użycie dysków zewnętrznych do przenoszenia danych z urządzeń nie posiadających połączenia z siecią. Dysk zewnętrzny przeznaczony do tego celu nie może być użytkowany w inny sposób.
2. Kopie zapasowe
 - 2.1 Każdy użytkownik komputera jest odpowiedzialny za dostosowanie i stosowanie mechanizmu kopiowania danych, tzn. musi określić przestrzeń na komputerze, na którą będą kopiowane wybrane dane z jego komputera służbowego.
 - 2.2 Kopia zapasowa będzie robiona zgodnie z ustalonym harmonogramem i będzie zapisywana na serwerach Instytutu.
3. Na każdym komputerze przetwarzającym dane badawcze wymagana jest bieżąca wersja systemu operacyjnego wraz z systemem antywirusowym.

7. Dokumenty związane

7.1. Polityka Zarządzania Danymi Badawczymi

8. Załączniki

1. Zał. 1_Rejestr urządzeń badawczych_RDM;
2. Zał. 2_Rejestr zarchiwizowanych danych badawczych.

9. Tabela zmian

Wersja/data wydania przed zmianą	Wersja/data wydania po zmianie	Treść zmiany/ numer strony/ punkt	Autor