ДУ «Інститут патології хребта та суглобів
імені професора М.І. Ситенка
Національної академії медичних наук України»

Інструкція з використання Colectica for Excel 7.5 (Standard Edition)

Адаптована для застосування для розміщення
FAIR даних в інституціональному репозитарії
ДУ «Інститут патології хребта та суглобів
ім. проф. М.І. Ситенка НАМН України»

Харків 2025

Інструкція з використання Colectica for Excel 7.5 (Standard Edition).
Адаптована для застосування для розміщення FAIR даних в інституціональному репозитарії ДУ «Інститут патології хребта та суглобів ім. проф. М.І. Ситенка НАМН України». Харків, 2025. – 30 с.

Інструкція укладено відповідно Політики відкритої науки ДУ «Інститут патології хребта та суглобів ім. проф. М.І. Ситенка НАМН України»

Colectica for Excel — це надбудова до Microsoft Excel, яка дозволяє документувати набори даних, додаючи метадані (описову інформацію) відповідно до стандарту DDI (Data Documentation Initiative). Це корисно для дослідників, аналітиків і організацій, які хочуть зробити свої дані зрозумілими, прозорими та придатними для повторного використання. Безкоштовна версія (Standard Edition) має базовий функціонал, але її достатньо для основної документації, яку планують розмістити в інституціональному репозитарії. Ця інструкція допоможе вам розпочати роботу з Colectica for Excel 7.5.

Для повної відповідності стандарту DDI для розміщення у сертифікованих репозитаріях відкритого доступу необхідні професійні версії або інші спеціалізовані програми.

Інструкція також містить правила опису колекції, які містять відкриті дані в репозитарії та рекомендації щодо формування та перевірки ключових слів відповідно стандартів MeSH.

ЗМІСТ

Colectica for Excel та роль метаданих у біомедичних дослідженнях 4
1. Підготовка таблиці даних 5
2. Встановлення Colectica for Excel 7.5 8
3. Активація Colectica for Excel 7.5 (Standard Edition) 9
4. Початок роботи: Ініціалізація документації 10
5. Документування набору даних (Dataset Details)..... 11
6. Документування змінних (Variable Details) 15
7. Етичні аспекти та анонімізація 21
8. Генерація документації (кодової книги) 22
9. Додаткові поради..... 22
10. Приклад повної документації..... 23
11 Розміщення даних в репозитарії 24
12. Словники ключових слів. 26
13. Інструкція з формування ключових слів для метаданих
наукових даних за допомогою MeSH (Medical Subject Headings)..... 28
14. Використання MeSH on Demand для визначення ключових слів 30

Colectica for Excel
та роль метаданих у біомедичних дослідженнях

У біомедичних дослідженнях обробка, інтерпретація та повторне використання даних вимагають не лише якісного збору, а й ретельного документування. У фокусі сучасних наукових підходів стоїть не лише результат, а й прозорість усього дослідницького процесу, що передбачає супровід кожного набору даних відповідними метаданими. Метадані — це структурований опис, який пояснює походження, зміст, формат і умови використання змінних, що використовуються в дослідженні. Їх наявність є критично важливою для забезпечення **відтворюваності**, **доступності** та **якісного обміну інформацією** в рамках мультидисциплінарних дослідницьких команд.

Colectica for Excel — це спеціалізована надбудова до Microsoft Excel, що дозволяє описувати змінні у табличних наборах даних згідно з міжнародним стандартом **DDI (Data Documentation Initiative)**. Вона інтегрується безпосередньо у знайоме середовище Excel і надає можливість автоматично формувати структуру метаданих, створювати версії, додавати мітки, типи даних, кодові списки, визначення змінних та їх опис. Це особливо цінно у біомедичних проєктах, де дані часто мають складну структуру (клінічні виміри, шкали, біомаркери, образні та лабораторні результати), і без чіткої документації ризикують втратити наукову цінність.

Створення метаданих з використанням Colectica забезпечує відповідність даних принципам **FAIR (Findable, Accessible, Interoperable, Reusable)**, які активно впроваджуються в медичну науку, системи охорони здоров’я та біобанки. Застосування цього інструменту дає змогу формалізувати опис змінних, полегшити інтеграцію даних між проєктами, пришвидшити аналіз та спростити архівування у відповідних наукових репозиторіях.

Таким чином, **Colectica for Excel** — це ефективне рішення для стандартизованої документації біомедичних даних, що робить дослідження більш структурованими, прозорими та придатними до багаторазового використання у майбутньому.

1. Підготовка таблиці даних

Дані в таблиці Excel мають бути підготовлені за певними принципами, щоб бути зрозумілими, повторно використовуваними та придатними для архівування або публікації.

Необхідно забезпечити якісну структуру таблиці для автоматичного створення метаданих без помилок, дублювань або втрат інформації.

Структура таблиці:

- Дані розміщені на **одному** аркуші Excel, що містить лише **один** набір спостережень.
- Перший рядок таблиці містить **унікальні назви** змінних (не мітки!).
- Кожен рядок → одна спостережувана одиниця (пацієнт, вимірювання, анкетування тощо).

Назви змінних

- Назви змінних — бажано **латиницею**, **без пробілів**, без `#`, `/`, `()`, `!`, тощо.
- Назви є **зрозумілими, короткими та унікальними** (`Sex`, `Age`, `VAS1`, `IL6_pre`).
- Уникати дуже загальні назви: `Data1`, `Var2`, `Test3`

Якість даних

- Всі значення у стовпці одного типу (числа / текст / дата).
- Немає об'єднаних клітинок (merged cells).
- Всі формули перетворено на значення (Paste as Values).
- Значення, що відсутні, позначено як `NA` або залишено порожнім (але позначено однаково)

Однорідність стовпців

- Немає стовпців-коментарів, виносок або службових колонок.
- Усі змінні, які потребують опису, **мають власний стовпець**, а не кілька вкладених значень в одній клітинці.

Супровід

Підготовлено окремий файл `README.txt`, де вказано:

- Назву дослідження
- Дати збору
- Тип дослідження (когортне, РКД тощо)
- Короткий опис, для чого збирались дані
- За потреби — окремий файл `ValueLabels.xlsx` для змінних з кодуванням (1 = Чоловік, 2 = Жінка тощо)

Приклад структури ReadMe.txt

```
==========================================
README: Набір біомедичних даних
==========================================

1. Назва дослідження:
"Вплив методів знеболення на рівень запалення у пацієнтів після
ендопротезування кульшового суглоба"

2. Автор(и) та установа:
Іваськін О.В., Інститут ортопедії та біомеханіки НАНУ

3. Контактна особа:
Email@email

4. Дата створення набору:
2024-12-10

5. Версія набору:
1.0

6. Опис даних:
Набір містить клінічні та лабораторні дані пацієнтів, які пройшли
лікування з приводу остеоартриту.
Зібрано демографічні дані, рівні IL-6, CRP, оцінки болю за
шкалою VAS, дані до та після лікування.
Загалом 120 спостережень, 15 змінних.

7. Формат:
Файл Excel (.xlsx), перший рядок — назви змінних, кожен
рядок — один пацієнт.
```

```
8. Визначення відсутніх значень:
Позначено як "NA"

9. Зміст стовпців (деталі у PDF/Colectica):
- ID: Унікальний ідентифікатор пацієнта
- Sex: Стать (1 = Чоловік, 2 = Жінка)
- Age: Вік (роки)
- Hb_pre/after: рівень Hb до/після лікування (г/л)
- Treatment: Тип лікування (1=фармакологічне, 2=хірургічне)
   ...

10. Умови використання:
Дані доступні для дослідницьких цілей за умови цитування авторів.

11. Пов'язані документи:
metadata.pdf — структуровані метадані, згенеровані у
Colectica
codebook.xls — словник змінних (якщо є)
publication_doi.txt — посилання на пов'язану публікацію
(якщо доступно)

12. Коментарі:
У деяких пацієнтів відсутні значення _____ через лабораторні
обмеження.
```

Готово до імпорту, якщо:

- Таблиця читається зверху вниз і зліва направо.
- Всі назви змінних унікальні.
- Немає форматування, що може вплинути на імпорт (злиття клітинок, приховані рядки).
- Є короткий супровід до таблиці — в окремому аркуші або файлі.

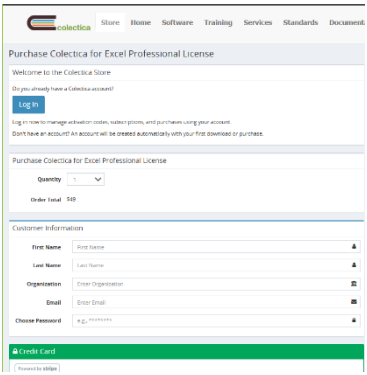
Рекомендовано зберігати файл у форматі .xlsx перед відкриттям у Colectica for Excel.

2. Встановлення Colectica for Excel 7.5

Системні вимоги

- Операційна система: Windows 10 та вище.
- Microsoft Excel: 2016 та вище, настільна версія Office 365 (macOS не підтримується).
- Інтернет: Для завантаження та активації.

Кроки встановлення



1. Перейдіть на офіційний сайт Colectica: (https://www.colectica.com).
2. У розділі Downloads знайдіть Colectica for Excel і завантажте версію 7.5 (Standard – безкоштовна).
3. Пройдіть реєстрацію.
4. На пошту прийде повідомлення з кодом активації та посиланням на завантаження інсталяційного файлу.
5. Запустіть завантажений файл (наприклад, `ColecticaForExcel-7.5.10650.exe`) і дотримуйтесь інструкцій інсталятора:

- Прийміть ліцензійну угоду.
- Виберіть папку для встановлення, або встановить за замовчуванням — (C:\Program Files\Colectica).
- Дочекайтеся завершення встановлення.

4. Відкрийте Microsoft Excel. Ви маєте побачити нову вкладку Colectica у стрічці меню.

Вирішення проблем
- Якщо вкладка Colectica не з’явилася:
- У Excel перейдіть до Файл > Параметри > Надбудови.
- У розділі Надбудови COM увімкніть Colectica for Excel.
- Перезапустить Excel.
- Якщо виникають помилки (наприклад, із DLL), перевстановіть програму або оновіть Excel.

3. Активація Colectica for Excel 7.5 (Standard Edition)

Безкоштовна версія потребує активації, інакше деякі функції (наприклад, створення списків кодів) можуть бути обмежені.

Кроки активації

1. Відкрийте Excel і перейдіть на вкладку ***Colectica.***

2. Натисніть ***Options*** (Налаштування).

3. У вікні, що відкриється, перевірте статус: якщо написано ***License: Colectica for Excel (Unregistered)***, програма не активована.

4. У поле ***Standard Activation Code*** введіть код активації, який прийшов на пошту.

8. Натисніть *Activate Standard* (Активувати Standard).

9. Натисніть **Save** (Зберегти) і перезапустить Excel.

10. Перевірте статус у ***Options***: має бути напис ***License: Colectica for Excel (Standard)***.

Примітка
- Якщо код не працює, перевірте підключення до інтернету або зверніться до служби підтримки Colectica через сайт.
- Активація потрібна лише один раз. Після цього всі функції Standard Edition будуть доступні.

4. Початок роботи: Ініціалізація документації

Підготовка даних

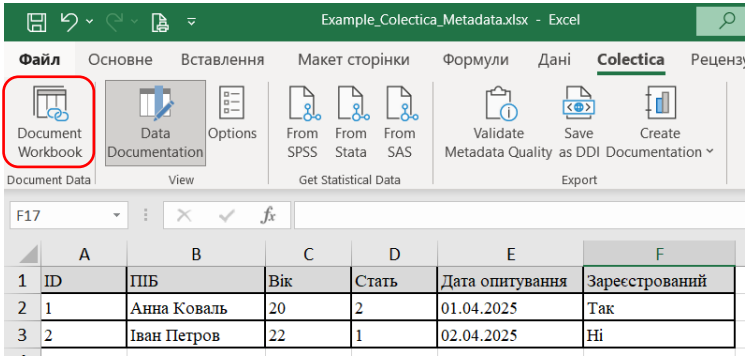
1. Відкрийте Excel-файл із вашим набором даних. Наприклад, у вас є таблиця:
2.

ID	ПІБ	Вік	Стать	Дата	Extensio	0	Flexio
1	Анна Коваль	20	2	01.04.2025	0	0	45
2	Іван Петров	22	1	02.04.2025	0	10	65

- Стовпці — це змінні (наприклад, Вік, Стать).
- Рядки — це записи (наприклад, респонденти).

Ініціалізація

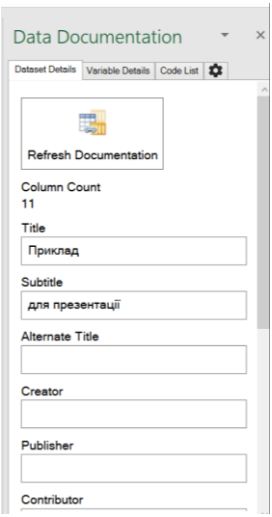
1. На вкладці Colectica у стрічці Excel натисніть ***Document Workbook*** (Документувати книгу).



2. У правій частині екрана з’явиться панель *Data Documentation* із вкладками:

- ***Dataset Details*** (Деталі набору даних).
- ***Variable Details*** (Деталі змінних).
- ***Code List*** (Список кодів).

5. Документування набору даних (Dataset Details)



Вкладка **Dataset Details** призначена для опису загальних метаданих про набір даних.

Перед описом набору даних та після будь-яких маніпуляції з ними (додавання рядків або змінних) треба перезапустити DataSet. Натисніть кнопку ***Refresh Documentation***.

У вкладці ***Dataset Details*** панелі ***Data Documentation*** у безкоштовній версії Colectica for Excel 7.5 доступні лише базові поля для опису набору даних. Нижче наведено повний перелік полів, доступних у Standard Edition, із розшифровкою їхнього призначення.

Поле	Розшифровка та призначення
Title *Назва*	**Назва набору даних.** Вказує офіційну або описову назву, яка ідентифікує набір даних. Наприклад, Опитування про зайнятість 2025. **Обов'язкове поле.** Використовується в кодових книгах, експорті DDI та для пошуку в репозиторіях. Допомагає швидко зрозуміти зміст даних.
Subtitle *Додаткова назва*	Доповнює основну назву (**Title**), надаючи додаткову інформацію, яка уточнює зміст, контекст або специфіку набору даних. Підзаголовок зазвичай використовується для короткого опису, який не вміщується в основну назву, але є важливим для розуміння. *Використовують* для уточнення географічного, часового або тематичного охоплення, для вказівки підпроєкту чи частини більшого дослідження, для додавання ключового аспекту, який не відображений у **Title**.
Alternate title	Надає іншу назву набору даних, яка може використовуватися в різних контекстах. Це може бути скорочена назва, абревіатура, переклад іншою мовою або назва, під якою набір даних відомий в іншому середовищі. Використовують якщо набір даних має офіційну назву, але

Поле	Розшифровка та призначення
	також відома неофіційна чи скорочена назва, набір даних має переклад назви іншою мовою (наприклад, англійською для міжнародного використання), набір даних має історичну назву або назву проєкту, яка відрізняється від офіційної.
Creator *Автор або організація*	Вказує особу чи організацію, яка створила або зібрала дані. Наприклад, ДУ Інститут патології хребта та суглобів…. Використовується для цитування та визначення авторства. Наприклад, ROR https://ror.org/02crmrs33.
Publisher *Відповідальний за публікацію*	Для інституціонального репозитарію вказується підрозділ, який розмістив набір даних. *Вписувати:* Назву підрозділу, який відповідає за публікацію, наприклад, Лабораторія …….
Contributor *Учасник*	Призначене для вказівки осіб, організацій або груп, які зробили внесок у створення, обробку чи поширення набору даних, але **не є основними авторами** (основні автори вказуються у полі **Creator**). Може включати: особи (дослідники, які допомагали зі збором даних), організації (інституції, які фінансували проект або надавали технічну підтримку).
Identifiers *Ідентифікатори*	Призначений для додавання унікальних ідентифікаторів набору даних, використовується для вказівки кодів або номерів, що ідентифікують набір даних у різних системах (наприклад, у репозиторіях, архівах чи базах даних). В Standart версії ідентифікатори додаються через вибір ⊞ **Приклади ідентифікаторів**: · Унікальний код проекту: xxxxUxxxxx · Внутрішній ідентифікатор організації: ЦФюxxx.xxxx. · DOI (Digital Object Identifier): 10.1234/stu.2025.001 (*в інституційному репозитарії DOI набору даних може супроводжувати публікацію)*
Copyright *Авторське право*	Вказує інформацію про авторські права на набір даних, зокрема власника прав, рік і умови використання. *Вписувати:* наприклад, © ДУ ІПХС ім. М.І. Ситенка…. або © 2025, ліцензія CC BY 4.0.
Access Rights *Права доступу*	Вказує умови доступу до набору даних, наприклад, хто може використовувати дані, чи є обмеження, чи потрібен дозвіл. *Вписувати*: Наприклад, Відкритий доступ для всіх користувачів. або Доступ обмежений, потрібен запит до ...

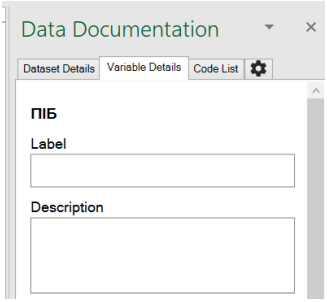
Поле	Розшифровка та призначення
В стандартній версії Colectica набор полів опису даних обмежений, і не повністю відповідає стандарту DDI. При необхідності відповідні поля можна додати вручну. *Для цього:* - Прокрутить донизу вкладки Dataset Details. - Натисніть ➕ (або Add). - Введіть назву поля, наприклад, Funding (Фінансування), і значення, наприклад, МОЗ. - Повторіть для інших полів, *Нижче показані поля, які рекомендовані додати до опису даних*. Власні поля дозволяють додавати будь-яку інформацію, наприклад, Region (Регіон) = Харків.	
Description *Опис набору даних.*	Дозволяє детально описати мету, зміст, методологію чи особливості даних. Наприклад, Дані зібрані шляхом опитування 1500 осіб у 2025 році. Допомагає користувачам оцінити придатність даних.
Date *Дата створення або збору*	Вказує дату, пов'язану з набором даних (створення, збір, публікація). Наприклад, 2025-04-01. Використовується для відстеження актуальності та версії. Рекомендується формат ISO (РРРР-ММ-ДД).
Subject *Тема або ключові слова.*	Вказує тематику даних через ключові слова. Наприклад, зайнятість, соціологія, Україна. Полегшує пошук і класифікацію в репозиторіях.
Language *Мова даних або метаданих*	Вказує основну мову, якою задокументовано дані чи метадані. Наприклад, Українська, англійська.
Coverage *Покриття*	Описує географічне, часове або тематичне покриття набору даних. Приклад: Географічне: Україна, всі регіони. Часове: Січень–Березень 2025. Тематичне: Доросле населення віком 18–65 років. Важливість: Допомагає користувачам оцінити, чи відповідає набір даних їхнім потребам (наприклад, чи покриває потрібний регіон чи період).
Source *Джерело*	Вказує джерело даних, якщо вони не зібрані безпосередньо автором (наприклад, інший набір даних, база даних, опитування). Допомагає відстежити походження даних і оцінити їх достовірність.

Поле	Розшифровка та призначення
	Приклад: Опитування, проведене компанією Research Ukraine. **Порада:** Якщо джерело складне, додайте деталі в **Description**
Version *Версія*	Вказує версію набору даних, якщо він оновлювався. Дозволяє відстежувати зміни в наборі даних і уникати плутанини між версіями Приклад: 1.0 або Оновлено 2025-04-10.

Заповнення полів опису набору даних є важливою передумовою для забезпечення його інтерпретованості, повторного використання та інтеграції у відкриті наукові системи. Навіть якщо деякі поля залишаються порожніми, старанне і максимально повне заповнення доступних елементів метаданих підвищує якість документації та сприяє цілісному розумінню змісту й призначення набору даних як для внутрішніх користувачів, так і для зовнішніх дослідників.

6. Документування змінних (Variable Details)

Вкладка **Variable Details** дозволяє описати кожну змінну (стовпець) в таблиці.



1. У панелі Data Documentation виберіть вкладку **Variable Details.**

2. У Excel виберіть стовпець, який хочете задокументувати (наприклад, стовпець ПІБ).

3. Заповніть поля:

Поле	Розшифровка та призначення
Label *Опис*	Наприклад, Прізвище, Ім'я По-батькові Інтерлекін IL-6
Description *Детальний опис змінної*	дозволяє надати детальний опис змінної, щоб пояснити її зміст, призначення, спосіб збору даних або інші важливі деталі. Це допомагає користувачам даних зрозуміти, що саме представляє змінна. *Вписувати:* Короткий опис змінної, який включає: · що вимірює змінна (наприклад, "Вік респондента на момент опитування"). · як дані були зібрані (наприклад, "Дані зібрані шляхом анкетування"). · Додаткові пояснення (наприклад, "Значення 99 означає пропущені дані"). · Опис має бути чітким і корисним для тих, хто використовуватиме набір даних. *Приклади: Для* **змінної "Age" (Вік).** "Вік респондента на момент опитування у роках. Дані отримані шляхом самозвітів. Значення 99 вказує на пропущені дані."
Data Tupe *Тип даних*	**Text** (*Текст*) – для текстових даних **Numeric** (*Числовий*) – Для чисел **DateTime** (*Дата/Час*) – Для дат **Code** (*Код*) – для категорій (наприклад, Стать = 1, 2). При введені даної опції стає доступна вкладка Code List для вводу кодів

Поле	Розшифровка та призначення
Type *Тип*	Type буде визначатися Data Type.

Numeric (*Числовий*). Використовується для змінних, які містять числові значення, придатні для математичних операцій (наприклад, додавання, обчислення середнього).

Підтипи:

- **BigInteger** — це числовий тип для дуже великих цілих чисел (10^{20})
- **Integer** (Ціле число):
  Наприклад, "Вік" = 20, 21.
- **Long** (Довге ціле):
  Для великих чисел, ID = 1234567890.
- **Short** (Коротке ціле):
  Для малих чисел, оцінки = 1, 2, 3.
- **Decimal** (Десяткове):
  Для дробових чисел, середній бал = 3.75.
- **Float** (Плаваюча кома):
  Для дробових чисел із середньою точністю, температура = 36.6.
- **Double** (Подвійна точність):
  Для дробових чисел із високою точністю, наприклад, 3.14159.
- **Count** (Підрахунок):
  Для невід'ємних чисел, наприклад, відвідування = 1, 10
- **Interval** (Інтервал):
  Для чисел із рівними інтервалами, температура = 0, 10, 20.

Обмеження: У безкоштовній версії не можна вказати діапазон значень (наприклад, min/max), що доступно в Professional Edition.

DateTime (*Дата/Час*). Для форматування цього типу даних використовують загально прийняті формати – дата, Час, день, місяць, рік, рік/місць і т.д.

Наприклад:

Date (Дата): 2025-04-12.

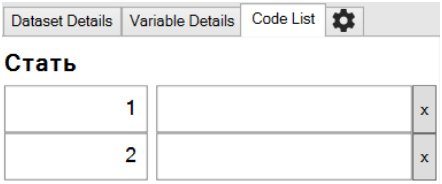
Time (Час): 14:30:00.

DateTime (Дата і час): 2025-04-12 14:30:00.

Розглянемо два параметри, які викликають багато запитань

Duration (Тривалість). Опис: Вказує тривалість часу у форматі ISO 8601 (наприклад, "P1Y2M3DT4H5M6S" — 1 рік, 2 місяці, 3 дні, 4 години, 5 хвилин, 6 секунд). **Використовують** для тривалості подій чи процесів, наприклад, час лікування, тривалість експерименту.

Приклад у медицині: Змінна "Treatment Duration" (Тривалість лікування) = "P3M" (3 місяці) → **Duration**.

Поле	Розшифровка та призначення
	Timespan (Часовий проміжок). Вказує період між двома моментами часу (початок і кінець), наприклад, "2025-04-01T10:00:00/2025-04-01T12:00:00". **Використовують**: Для фіксації часового діапазону, наприклад, період госпіталізації, час спостереження. **Приклад у медицині**: Змінна "Hospital Stay" (Період госпіталізації) = "2025-04-01/2025-04-05" → **Timespan**. **Code** відкриває вкладку Code List (Create New Code List створення нового листа кодів, чи Use existing code list – вибір вже існуючого набору кодів) **Code → Create new code list** створює у вкладці Code List таблицю для вводу  Опція **Use existing code list** дозволяє вибрати вже створений список кодів, який був визначений раніше у вашому проекті, замість створення нового списку для цієї змінної. *Наприклад, Код інтенсивності болю впродовж декількох етапів спостереження, закодовані за однією схемою*. Це корисно, коли однакові коди застосовуються до кількох змінних (наприклад, "Стать" і "Регіон" можуть мати однакові списки кодів у різних наборах даних).
Minimum / Maximum	Поля дозволяють задокументувати діапазон допустимих значень для числових змінних (з типом даних **Numeric**) або, у деяких випадках, для інших типів, якщо це доречно. Ось короткий опис: **Minimum** (Мінімум): Вказує найменше допустиме значення для змінної. **Maximum** (Максимум): Вказує найбільше допустиме значення для змінної.

<table>
<tr><th>Поле</th><th>Розшифровка та призначення</th></tr>
<tr><td>Scale
Шкала</td><td>Визначає тип шкали вимірювання змінної, що вказує на її математичні властивості та можливі операції з даними.
Поле Scale у вкладці очікує числове значення
<table>
<tr><th>Числове значення</th><th>Тип шкали</th><th>Пояснення</th></tr>
<tr><td>0</td><td>Nominal</td><td>Номінальна шкала
(категорії без порядку)</td></tr>
<tr><td>1</td><td>Ordinal</td><td>Порядкова шкала
(є порядок, без інтервалів)</td></tr>
<tr><td>2</td><td>Interval</td><td>Інтервальна шкала
(є порядок і інтервали, але немає абсолютного нуля)</td></tr>
<tr><td>3</td><td>Ratio</td><td>Шкала відношень (є порядок, інтервали і абсолютний нуль)</td></tr>
</table>
Приклади визначення типу шкал для даних
<table>
<tr><th>Назва змінної</th><th>Значення</th><th>Scale</th></tr>
<tr><td>Sex</td><td>1 = Чоловік,
2 = Жінка</td><td>Nominal</td></tr>
<tr><td>TreatmentGroup</td><td>1 = Хірургічне,
2 = Консервативне</td><td>Nominal</td></tr>
<tr><td>Age</td><td>Вік у роках</td><td>Ratio</td></tr>
<tr><td>CRP</td><td>С-реактивний білок у мг/л</td><td>Ratio</td></tr>
<tr><td>Temperature_C</td><td>Температура у °C</td><td>Interval</td></tr>
<tr><td>DiseaseStage</td><td>Стадія хвороби
(I, II, III)</td><td>Ordinal</td></tr>
</table></td></tr>
<tr><td>Decimal Position
Десяткова позиція</td><td>Визначає кількість знаків після коми для числових змінних із дробовою частиною (з типом даних Numeric, підтипи Decimal, Float, Double).
Використовують для дробових чисел, щоб вказати точність:
Decimal Position = 1: 12.3 (1 знак після коми).
Decimal Position = 2: 12.34 (2 знаки після коми).
Приклади</td></tr>
</table>

<table>
<tr><th>Поле</th><th>Розшифровка та призначення</th></tr>
<tr><td></td><td>Змінна Temperature: 36.6 → Decimal Position = 1.
Змінна Blood Pressure: 120.50 → Decimal Position = 2.</td></tr>
<tr><td>Variable Weight
Вага змінної</td><td>Використовується для позначення змінної, яка є статистичною вагою у наборі даних. Ваги (weights) застосовуються у статистиці для коригування даних, наприклад, щоб врахувати нерівномірність вибірки чи зробити вибірку репрезентативною.</td></tr>
<tr><td>Additivity
Адитивність</td><td>Використовується для визначення характеру змінної у контексті часу або агрегації. Може приймате одне зі вказаних значень:
Unspecified означає, що тип адитивності не застосовується або не встановлений
Stock вказує на фіксацію показника у певний момент часу
Flow — на його накопичення або зміну за інтервал часу.. Це особливо корисно для аналізу часових рядів або динаміки клінічних показників."
Non-additive (Неадитивна): – значення не можна підсумовувати або усереднювати, оскільки така операція не має змісту або спотворює дані.
Приклад визначення:
<table>
<tr><th>Змінна</th><th>Additivity</th><th>Пояснення</th></tr>
<tr><td>Рівень глюкози в крові</td><td>Stock</td><td>Це значення на момент часу (напр., натще)</td></tr>
<tr><td>Об’єм сечі за добу</td><td>Flow</td><td>Агрегується за часовий інтервал</td></tr>
<tr><td>Загальна доза препарату</td><td>Flow</td><td>Розраховується як сума доз за дні лікування</td></tr>
<tr><td>Температура тіла</td><td>Stock</td><td>Становий показник — фіксується в момент вимірювання</td></tr>
<tr><td>Наявність діагнозу (0/1)</td><td>Unspecified</td><td>Категоріальна змінна, не потребує типу адитивності</td></tr>
</table></td></tr>
<tr><td>Measurement Unit
Одиниця виміру</td><td>Вказує одиницю, в якій вимірюється значення змінної. Документує одиницю виміру для змінної, щоб уточнити, як інтерпретувати значення. Застосовується до числових (Numeric) або інших змінних, де одиниця виміру має значення.</td></tr>
</table>

Поле	Розшифровка та призначення
	Приклади: Змінна Вік: **Measurement Unit** = роки. Змінна Тиск: **Measurement Unit** = мм рт. ст. Змінна Зріст: **Measurement Unit** = сантиметри.
Role *Роль*	Вказує, яку роль виконує змінна у наборі даних, наприклад, ідентифікатор, вага, залежна змінна тощо. Допомагає користувачам даних зрозуміти призначення змінної. В Standart версії має два показника: **Identifier** (Ідентифікатор): Змінна є унікальним ідентифікатором (наприклад, "Patient ID"). **Weight** (Вага): Змінна використовується як статистична вага (наприклад, "Sample Weight").
Analysis Unit Одиниця аналізу	Вказує, до якої одиниці відносяться дані змінної. Має доволі багато варіантів відповіді: **Individual** (Індивід): Дані про особу (наприклад, вік). **Household** (Домогосподарство): Дані про домогосподарство (наприклад, дохід сім'ї). **Organization** (Організація): Дані про організацію (наприклад, бюджет лікарні). **Event** (Подія): Дані про подію (наприклад, дата візиту). **Geographic Unit** (Географічна одиниця): Дані про географічну одиницю (наприклад, регіон). **Not specified** (Не вказано): Одиниця не визначена. *При необхідності можна додати свою категорію.*
Response Unit Одиниця відповіді	Вказує, хто або що є джерелом даних для змінної. Має варіанти: **Individual** (Індивід): Дані надані окремою особою (наприклад, "Age" = вік, вказаний самим респондентом). **Household** (Домогосподарство): Дані надані від імені домогосподарства (наприклад, "Household Size" = розмір сім'ї, вказаний головою сім'ї). **Organization** (Організація): Дані надані організацією (наприклад, "Hospital Staff Count" = кількість персоналу, вказана адміністрацією). **Proxy** (Представник): Дані надані представником (наприклад, "Child Health" = здоров'я дитини, вказане батьками).

Поле	Розшифровка та призначення
	Not specified (Не вказано): Джерело даних не визначене. *При необхідності можна додати свою категорію*

Response Unit та **Analysis Unit** мають різні призначення, хоча й можуть здаватися схожими. Ось їхня відмінність:

Response Unit вказує, хто або що є джерелом даних для змінної, тобто хто надав інформацію. *Наприклад:* Для змінної **Child Age** (Вік дитини) Response Unit = **Proxy** (батьки надали дані про дитину).

Analysis Unit вказує, до кого або чого відносяться дані змінної, тобто об'єкт аналізу. *Наприклад:* Для тієї ж змінної **Child Age,** Analysis Unit = = **Individual** (дані стосуються окремої дитини

Нижче показані поля, які можна додати власні поля через "+" (Custom Fields).

Missing Values (Пропущені значення): Вказує, які значення означають пропуски (наприклад, "99").

Source (Джерело): Описує джерело даних (наприклад, "Анкета").

Collection Method (Метод збору): Як зібрані дані (наприклад, "Опитування").

Precision (Точність): Точність вимірювання (наприклад, "±0.1").

Frequency (Частота): Частота збору даних (наприклад, "Щомісяця").

Notes (Примітки): Додаткові коментарі (наприклад, "Дані перевірені").

7. Етичні аспекти та анонімізація

Якщо у книзі є лист із повними даними (для роботи) і лист із анонімізованими даними (для відкритого доступу), краще створити окремий файл (наприклад, **Public_Data.xlsx**) лише з анонімізованим листом для публікації.

Файл з повними даними (наприклад, **Full_Data.xlsx**) краще документувати окремо для внутрішнього використання, а анонімізований — для відкритого доступу з **Access Rights** = Відкритий доступ.

Якщо залишаєте один файл, у **Dataset Detail,** треба вказати:

"Sheet 1 Description" = "Повні дані, доступ обмежений",

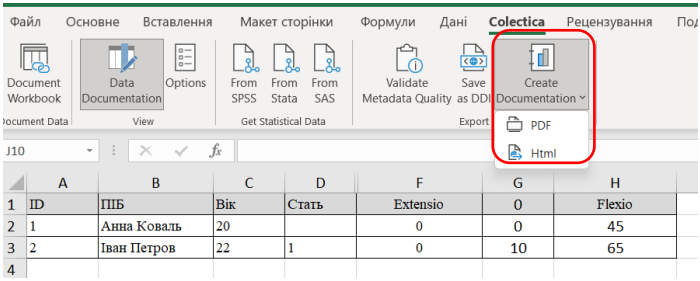
"Sheet 2 Description" = "Анонімізовані дані, відкритий доступ"

Увага! Переконайтеся, що повні дані не будуть опубліковані.

8. Генерація документації (кодової книги)

Після документування набору даних і змінних ви можна перейти до створення кодової книги — документ, який містить усі метадані.

1. На вкладці Colectica у стрічці Excel натисніть **Create Documentation** (Створити документацію).



2. Виберіть формат:

PDF: Основний формат, доступний у Standard Edition.

HTML: Дозволяє переглядати документацію у браузері.

3. Оберіть маршрут збереження файлу.

4. Натисніть **Зберегти**.

5. Збережіть файл на вашому комп'ютері (наприклад, biomech_2025.pdf або biomech_2025.html) і відкрийте його, щоб переглянути документацію.

* Export метаданих у формат DDI (XML) для використання в інших програмах для Standart версії не розглядаємо, через складну процедуру форматування .xml файла вручну. Для цього необхідні додаткові програми, або функціонал Zenodo, figShare, OSF та ін.

9. Додаткові поради

Організація даних: Перед початком переконайтеся, що ваш Excel-файл структурований (стовпці — змінні, рядки — записи).

Регулярне збереження: Зберігайте файл Excel після кожного етапу, щоб уникнути втрати метаданих.

Перевірка: Генеруйте тестову кодову книгу після кожного етапу, щоб переконатися, що метадані відображаються коректне.

10. Приклад повної документації

Knee_vibration

Title	Knee_vibration
Subtitle	Вібраційна розробка колінного суглоба після іммобілізації внаслідок позасуглобових переломів кісток нижньої кінцівки
Alternate title	Вібраційний вплив частотою 20 Гц
Copyright	ROR https://ror.org/02crmrs33.
Access rights	Open
Identifier	№ держреєстрації 0121U111555
Identifier	ЦФ 2020.2
Variable Count	13

Id - Ідентифікатор пацієнта

Type	Numeric (Integer)
Numeric Details	Decimals: 0, Scale: 0

Day - День процедури

Type	Numeric (Integer)
Numeric Details	Decimals: 0, Scale: 1
Description	Процедури проводили щодня, впродовж 2 тижнів. Для аналізу взяти 1,2,5, 10 та 14 сеанси

Іммоб - Термін іммобілізації

Type	Numeric (Count)
Measurement Unit	місяців
Numeric Details	Decimals: 0, Scale: 3
Description	Термін іммобілізації кінцівки до початку проведення процедур у місяцях
Analysis Unit	Individual

Ext_0 - Кут розгинання коліна до процедури

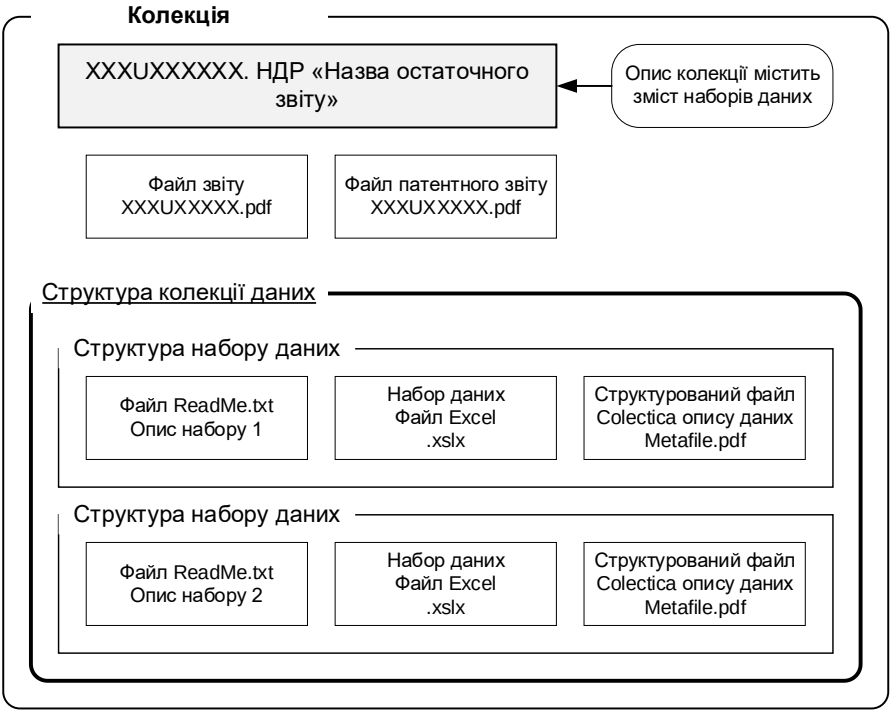
Type	Numeric (Decimal)
Measurement Unit	градуси
Numeric Details	Decimals: 0, Scale: 3
Analysis Unit	Individual

Flex_0 - Кут згинання до процедури

Type	Numeric (Decimal)
Measurement Unit	градуси
Numeric Details	Decimals: 0, Scale: 3

11 Розміщення даних в репозитарії

Структура розміщення файлів звітів в інституціональному репозитарії



У репозитарії ДУ «Інститут патології хребта та суглобів ім. проф. М.І. Ситенка НАМН України» розміщення електронних документів здійснюється відповідно до затвердженої структури.

У фонді **«Звіти НДР»** для кожного звіту створюється **окрема колекція** з назвою за шаблоном:

«№ держреєстрації» + «НДР. Назва» + «період виконання»

У описі колекції розміщуються загальні відомості про науково-дослідну роботу, а також структурований перелік наявних наборів даних (колекцій даних) за наступним шаблоном:

Опис <u>колекції</u> для розміщення в репозитарії DSpace:

Відомості про НДР у форматі:

Номер державної реєстрації «*<u>повна назва теми НДР»</u>* (дата початку – дата завершення).

Приклад: 0119U102343 Розробити комплексний підхід до лікування уродженого псевдоартрозу кісток гомілки у дітей (2020-2022)

Колекція містить:

- файл звіту про НДР;
- супровідний патентний звіт (за наявності);
- структурований опис набіру (-ів) даних

Наявні набори даних:

1. Набір 1 – короткий опис: що містить, кількість змінних, період збору
2. Набір 2 – аналогічний опис

Кожен набір даних супроводжується:

- файлом readme_назва.txt з описом структури;
- файлом meta_назва.pdf з метаданими (створений у Colectica);

Повідомлення: *Доступ до матеріалів надається відповідно до політики відкритої науки. У разі використання даних, посилатися на колекцію репозитарію.*

До складу кожної колекції входять:

- файл звіту (.pdf). Назва складається з № держ реєстрації
- файл патентного звіту (за наявності) (.pdf). Назва «Пат_звіт_ XXXXUXXXX»
- один або кілька **наборів даних** із супровідною документацією:
    - файл набору даних **`*.xlsx`**
    - файл **`readme.txt`** з описом змісту,
    - файл **`meta_*.pdf`** з метаданими, згенерований Colectica або подібного інструменту.

У межах набору даних, **використовувати узгоджене іменування файлів**, що включає унікальний ідентифікатор/назву набору (наприклад, biochem_blood) та префікс типу супровідного документа (readme_, meta_).

Наприклад: biochem_blood.xlsx; readme_ biochem_blood.txt, meta_ biochem_blood.pdf.

Такий підхід забезпечує прозорість та впорядкованість архіву, зручність доступу для користувачів, відповідність вимогам відкритої науки та принципів FAIR

12. Словники ключових слів

Ключові слова (keywords) у науковій документації мають не просто описовий характер – вони мають спиратися на стандартизовані словники або тезауруси, особливо в медицині. Це забезпечує індексацію, інтероперабельність і пошук, а також відповідність міжнародним стандартам відкритої науки.

Найбільш вживані медичні словники ключових слів

*1. **MeSH (Medical Subject Headings).***

Офіційний тезаурус Національної медичної бібліотеки США (NLM). Широко використовується в PubMed, Medline, Cochrane, Scopus.

Це контрольований словник (тезаурус), який розроблено Національною медичною бібліотекою США (NLM) для стандартизації медичних термінів. Використовується в базах PubMed, Medline, а також у системах відкритого

Доступний українською (обмежено), англійською повністю
https://meshb.nlm.nih.gov/search

*2. **SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms)***

Клінічна онтологія, що використовується в електронних медичних записах (EHR). Розширена, ієрархічна структура, підходить для систематизації даних. **Зазвичай використовується в клінічних системах, не для публікацій**, але якщо дані підлягають стандартизації — може бути джерелом для кодів.
https://browser.ihtsdotools.org/

*3. **ICD-10 / МКХ-10 (Міжнародна класифікація хвороб)***

Для клінічних діагнозів, підходить для супровідних змінних у медичних наборах

https://icd.who.int/browse10/2019/en
Українська: http://medstat.gov.ua

*4. **UMLS (Unified Medical Language System)***

Об’єднує MeSH, SNOMED, LOINC, ICD, ATC та інші системи. Потребує реєстрації

https://www.nlm.nih.gov/research/umls/index.html

Вибір словника

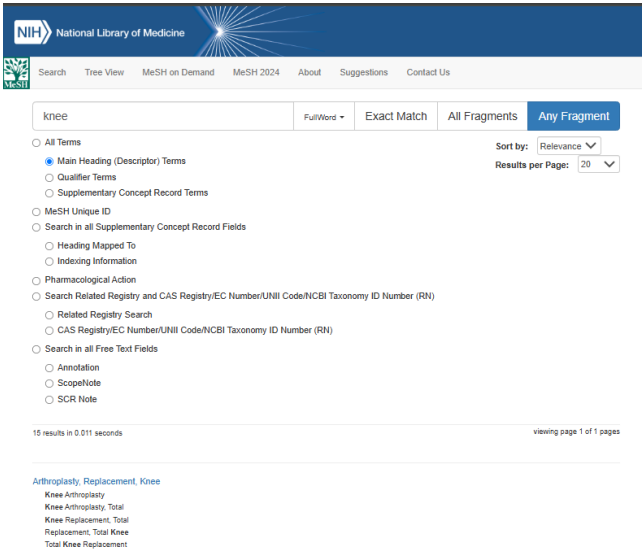
Мета	Рекомендований словник
Публікація, репозитарій, пошук	**MeSH**
Діагностичні коди	**ICD-10**
Електронна історія хвороби	**SNOMED CT**
Машинозчитувана відповідність	**UMLS** (складний, але потужний)

Рекомендація для політики репозитарію:

Ключові слова до колекцій, звітів і наборів даних мають відповідати контрольованим словникам, рекомендованим у медичній спільноті, зокрема MeSH (Medical Subject Headings). Це дозволяє забезпечити правильну індексацію, повторне використання та пошук.

13. Інструкція з формування ключових слів для метаданих наукових даних за допомогою MeSH (Medical Subject Headings)

MeSH (Medical Subject Headings) —доступу для полегшення пошуку та індексації.



Крок 1. Перейдіть до офіційного браузера MeSH

https://meshb.nlm.nih.gov/search

Крок 2. У рядку пошуку введіть ключове слово англійською мовою

Наприклад: "osteoarthritis", "inflammation", "biomarkers"

Якщо ви не знаєте точного терміна, використовуйте синоніми або описові слова або можна скористатися MeSH on Demand (посилання нижче).

Крок 3. Оберіть відповідний результат у списку

Знайдіть термін, позначений як **Main Heading (Descriptor)** – основний термін MeSH

Крок 4. Перегляньте додаткову інформацію

Scope Note — визначення терміна, що допомагає з інтерпретацією

Entry Terms — список синонімів і альтернативних назв, які допоможуть уникнути дублювання

Крок 5. Повторіть пошук для кожного ключового поняття теми дослідження

Визначте терміни для захворювання, анатомічної області, процесу, втручання, методів оцінки тощо

Крок 6. Відібрані MeSH-терміни запишіть у поле "Ключові слова"

Використовуйте **англійські варіанти MeSH-термінів**

Вказуйте **3–8 ключових слів**, розділених комами

Починайте з загальних термінів, потім вужчі й методичні

Приклади пошуку:

Пошуковий запит	**Обраний mesh-термін**	**Додаткова інформація**
Osteoarthritis	Osteoarthritis	Хронічне захворювання суглобів
Knee	Knee Joint	Анатомічна структура
Injury knee	Knee Injuries	Ушкодження менісків, тощо
Risk	Risk Factors	Фактори розвитку захворювань
Instability	Joint Instability	Порушення стабільності суглоба
Cartilage	Cartilage	Хрящова тканина
Arthroscopy	Arthroscopy	Метод візуалізації і лікування

УВАГА!

Уникайте абревіатур, якщо немає офіційного MeSH-відповідника

Не використовуйте занадто вузькі неіндексовані терміни

Використовуйте однакові терміни для подібних робіт в одному фонді

Додаткові ресурси

Браузер MeSH: https://meshb.nlm.nih.gov/search

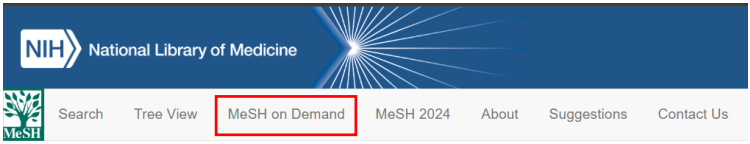
MeSH on Demand: https://meshb.nlm.nih.gov/MeSHonDemand

Довідка по MeSH: https://www.nlm.nih.gov/mesh/meshhome.html

Ця інструкція рекомендована для використання при заповненні метаданих у репозитарії, а також у супровідних файлах README і метаданих до наборів даних.

14. Використання MeSH on Demand для визначення ключових слів

Крок 1. Перейдіть MeSH on Demand
https://meshb.nlm.nih.gov/MeSHonDemand
або в меню виберіть необхідну опцію



Крок 2. У вікно помістити назву теми або опис набору даних. Натиснути Search.

Програма виділить ключові терміни, відповідно MeSH. Вони будуть розміщені як перелік у правому вікні.

Нижче буде показано декілька посилань, які знайдені в PubMed за комбінацією даних ключових слів.

За необхідністю можна виконати розширений пошук, натиснувши на кнопку «***Start PubMed Search***».

Зберегти перелік ключових слів можна натиснувши кнопку «***Export Data***»

Функції MeSH on Demand:

- Виділення в поданому тексті термінів, які відповідають лексиці MeSH.
- Швидкий перегляд приміток щодо терміну дескриптора MeSH.
- Інтерфейс для пошуку в PubMed за допомогою словника MeSH, знайденого MeSH on Demand, і відповідних ширших термінів.
- Функція експорту даних MeSH для збереження даних MeSH on Demand у текстовий файл.