

FAIR ДАНІ ТА УПРАВЛІННЯ ДОСЛІДНИЦЬКИМИ ДАНИМИ

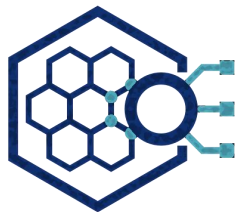
Навчальний курс -2026

Лідія Гладченко

к.е.н., доцент

Лабораторія екосистем відкритих інновацій, КАУ





RESEARCH DATA - OPEN BY DEFAULT



FAIR ДАНИ

FAIR ДАНІ

FAIR дані - це структуровані, документовані, ліцензовані цифрові дані, які організовані та керовані у такий спосіб, що **максимізує їхню наукову цінність** та можливість використання як людьми, так і комп'ютерними системами.

FAIR дані –цифрові дані, що відповідають принципам *знаходжунваності, доступності, інтероперабельності та повторного використання:*

- FAIR дані легко знаходяться, їх можна багаторазово використовувати
- FAIR дані сприяють вищому рівню цитованості
- FAIR дані є вимогою грантодавців і журналів
- FAIR дані відповідають концепції відкритої науки і забезпечують відтворюваність досліджень і довіру до їх результатів

FAIR дані не тотожні відкритим даним

- **Відкриті дані (Open Data)** – фокус уваги на легкому доступі
- **FAIR дані** - дані, які організовані так, щоб їх **можна було знайти, зрозуміти й повторно використати**, незалежно від рівня відкритості

Принцип відкритих даних: «as open as possible, as closed as necessary»

Довідково:

Відкриті дані – оприлюднені у цифровому вигляді дані, отримані в ході дослідження, з можливістю будь-якому користувачеві їх завантажувати, копіювати, повторно обробляти для будь-яких цілей без фінансових, юридичних або технічних перешкод

ПРИНЦИПИ FAIR

F – A – I – R = Findable – Accessible – Interoperable – Reusable

Дотримання принципів FAIR потрібне на етапі поширення даних

Findability – відшукуваність даних – легко знайти людям та комп'ютерним системам

Метаданим присвоюється глобально унікальний ідентифікатор (DOI)

Дані описуються за допомогою детальних та повних метаданих

Метадані реєструються або індексуються в пошуковому ресурсі

Метадані визначають ідентифікатор даних

Accessibility – доступність, збереження та можливість завантаження.

Метадані можна знайти за їх ідентифікатором за допомогою стандартизованого протоколу зв'язку

Протокол є відкритим, безкоштовним та універсально реалізованим

Протокол передбачає процедуру автентифікації та авторизації, де це необхідно

Метадані є доступними, навіть коли дані більше не є доступними

ПРИНЦИПИ FAIR

Interoperability – інтероперабельність – дані готові до поєднання з іншими пакетами даних / інтеграції до інших досліджень

- Метадані використовують формальну, доступну, спільну та широко застосовувану мову для представлення знань

- Метадані використовують словники, які відповідають принципам FAIR

- Метадані містять кваліфіковані посилання на інші метадані

Reusability – придатність до повторного використання - дані готові до використання в подальших дослідженнях та обробці з використанням обчислювальних методів

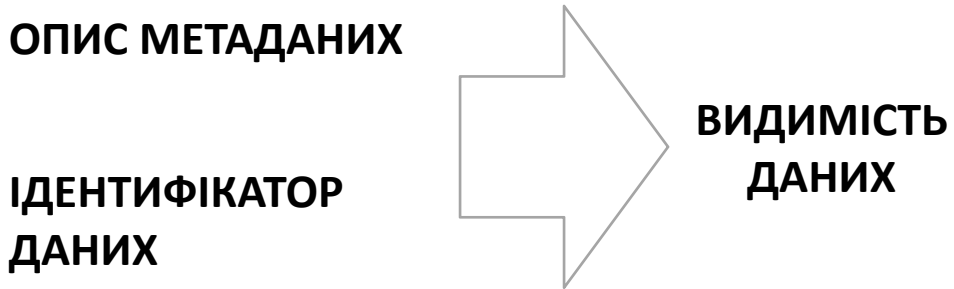
- Метадані можуть мати велику кількість точних і відповідних атрибутів

- Метадані оприлюднюються з чіткою та доступною ліцензією на використання даних

- Метадані описують умови отримання даних, експериментальну установку, методологію і т.д.

- Метадані відповідають стандартам наукової спільноти

Findability – відшукуваність даних



Ідентифікатор даних – унікальний буквенно-цифровий код, що однозначно визначає дані.

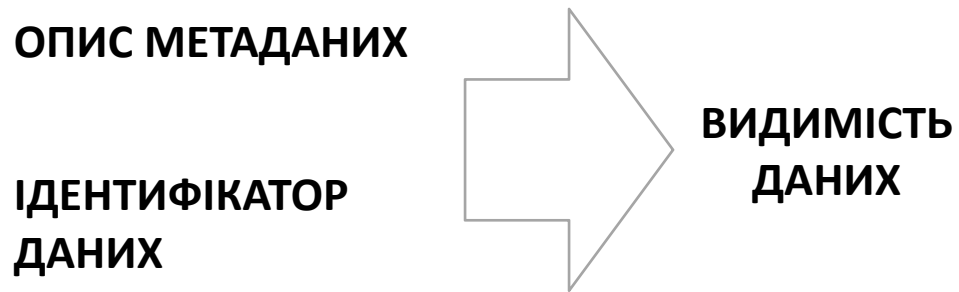
➤ Глобально унікальні ідентифікатори:

- DOI – Digital Object Identifier – забезпечує постійне посилання на опис набору даних, що дозволяє надійно знаходити і цитувати дані, навіть якщо змінюється їх фізичне розташування
- PURL – Persistent Uniform Resource Locator - діє через перенаправлення на поточне місцезнаходження ресурсу і дозволяє зберегти постійне посилання у випадку, якщо URL-адреса змінюється
- Handle – використовується для унікальної ідентифікації цифрових об'єктів.

➤ Веб-адреса (URL) – посилання на веб-сторінку, де було розміщено дані.

➤ Локальний ідентифікатор – діє в межах певної організації / системи, не є глобально унікальним.

Findability – відшукуваність даних



Метадані (Metadata Record) – структурована інформація про набір даних, яка описує його характеристики:

- Опис вмісту набору даних
- Інформацію про якість даних
- Формат даних
- Місцезнаходження даних
- Контактну інформацію особи/ організації, яка відповідальна / управляє даними
- Ідентифікатори пов'язаних людей, місць, речей (грантових конкурсів і т.п.)

Метадані є машиночитаною документацією даних

МЕТАДАНІ

Метадані можуть описувати

- Фізичні об'єкти
- Цифрові об'єкти

Збереження метаданих

- Реєстри – містять лише записи (опис) метаданих
- Репозитарії – містять метадані, а також забезпечують зберігання і доступ до самих даних.

Типи репозитаріїв:

- Загальні репозитарії (Zenodo)
- Галузеві / тематичні репозитарії
- Інституційні репозитарії (як правило, університетські)
- Репозитарії текстів і репозитарії даних

МЕТАДАНІ

МІНІМАЛЬНИЙ FAIR-НАБІР МЕТАДАНИХ

Поле метаданих	Чому це важливо для дослідника
Назва набору даних (Title)	Допомагає іншим правильно знайти та процитувати дані; без назви набір губиться серед інших і не ідентифікується у репозиторіях.
Автор / відповідальна особа (Creator / PI)	Забезпечує належну атрибуцію та можливість отримати контакт для уточнення даних; підвищує цитованість дослідника.
Короткий опис (Description)	Дозволяє іншим зрозуміти, що містять дані, без відкриття всіх файлів; підвищує шанси, що дані будуть використані та процитовані.
Ключові слова (Keywords)	Покращують знаходжуваність у каталогах; без них дані практично «невидимі».
Дата створення / період збору (Date Created / Collection Period)	Дає контекст для інтерпретації даних; дозволяє досліднику довести актуальність і відтворюваність набору.
Версія набору (Version)	Знімає ризики плутанини між різними редакціями; важливо, якщо дані оновлювалися або виправлялися.
Тип та формат даних (Data Type / Format)	Дозволяє іншим відкрити та використовувати дані; неправильний формат ускладнює аналіз або робить дані непрацездатними.

МЕТАДАНІ

МІНІМАЛЬНИЙ FAIR-НАБІР МЕТАДАНИХ

Поле метаданих	Чому це важливо для дослідника
Метод збору та обробки (Methodology)	Дає можливість відтворити дослідження та оцінити якість даних; без цього дані не вважаються надійними.
Список змінних (Variable List / Codebook)	Забезпечує правильну інтерпретацію даних; без словника змінних навіть якісний набір не можна адекватно використовувати.
Ліцензія (License)	Однозначно визначає, що дозволено робити з даними; захищає дослідника від порушення авторських прав і обмежує небажане використання.
Умови доступу (Access Rights)	Дає змогу контролювати, хто може бачити чи завантажувати дані; важливо для чутливих даних та дотримання законодавства.
Постійний ідентифікатор (DOI або інший PID)	Гарантує стабільне посилання та цитованість; без PID дані можуть зникнути або бути некоректно процитовані.
Версія набору (Version)	Знімає ризики плутанини між різними редакціями; важливо, якщо дані оновлювалися або виправлялися.
Місце збереження (Repository / Location)	Забезпечує довготривале збереження; дані в особистих хмарах або на робочому комп'ютері не відповідають FAIR і можуть бути втрачені.

Accessibility - доступ і довготривале зберігання даних

Визначення правил доступності відбувається на основі **плану управління даними**

Доступ до даних передбачає опис таких **ключових параметрів**:

- Хто може отримати доступ до даних
- За яких умов надається доступ (наприклад, для некомерційного використання)
- Коли може відбутись доступ. Встановлення строку, на який обмежується доступ до даних (період затримки), називається **ембарго** (причини полягають у питаннях безпеки, конфіденційності, власності, комерційного використання).

Управління доступом потребує визначення **ліцензії**, на умовах якої буде надаватись доступ до даних.

Доступність передбачає також **збереження запису метаданих (metadata record availability) навіть тоді, коли сам набір даних став недоступним**. Це історичний запис, який уможливорює цитування, відстеження історії даних, дозволяє знайти/ звернутись до дослідників/ авторів/ власників даних.

ЛІЦЕНЗІЇ

Ліцензії на дані – це юридичні рамки, що встановлюють дозволи та обмеження при роботі з даними і визначають, як набори даних можна використовувати, поширювати, змінювати

За будь-якої ліцензії використання/ поширення/ зміна даних має відбуватись на принципах **доброчесності**.

Отримання ліцензій забезпечують:

- Creative Commons (CC) – неприбуткова організація, яка розробила сімейство ліцензій на безкоштовній основі (інструмент Chooser: <https://creativecommons.org/chooser/>)
- Open Data Commons – Ліцензії, спеціально розроблені для баз даних
 - Open Database License (ODbL)
 - Public Domain Dedication and License (PDDL)
- Інші нестандартні ліцензії з унікальними умовами, у тому числі організаційними обмеженнями

Види ліцензій CC

- CC0 – Передає дані / результати досліджень у суспільне використання без обмежень
- CC BY – Дозволяє використання даних за належного посилання на авторів
- CC BY SA – Використання даних за належного посилання та поширення похідних робіт на тих самих умовах (share alike)
- CC BY-NC - Використання даних за належного посилання виключно на некомерційних засадах (noncommercial)

RECOMMENDED LICENSE



Creative Commons Attribution 4.0 International

This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, even for commercial purposes.

 **BY:** Credit must be given to you, the creator.

[See the License Deed](#) 

Найбільш широка
ліцензія, яка дозволяє
використовувати роботу
за умови належного
цитування

Interoperability - інтероперабельність

Інтероперабельність – готовність даних до обробки, поєднання з іншими пакетами даних та інтеграції до інших досліджень.

Основа інтероперабельності – використання **машиночитаних та відкритих форматів даних**, завдяки чому різні програми та системи можуть ефективно взаємодіяти з даними.

Машиночитані дані можуть **автоматично зчитуватись та оброблятись комп'ютерними системами** без людського втручання. Найбільш розповсюджені формати машиночитаних даних є:

- CSV
- JSON – JavaScript Object Notation
- XML – eXtensible Markup Language
- RDF – Resource Description Framework

Принцип Інтероперабельності потребує **спільного розуміння термінів і понять** між різними користувачами і системами, для чого використовуються словники.

Словники (Vocabularies) – структуровані набори термінів, які використовуються дослідниками/ командами для стандартизованого опису даних та їх змісту таким чином, щоб вони могли бути зрозумілі людям і машинам (списки термінів, глосарії, класифікатори, таксономії тощо).

ЗВ'ЯЗАНІ ДАНІ

Для реалізації принципу інтеоперабельності використовуються зв'язані дані (Linked Data)

Зв'язані дані (Linked Data) – дані, між якими при публікації в мережі інтернет створені змістовні зв'язки між різними наборами даних та ресурсами. Вони є машиночитаними, структурованими, доступними онлайн, базуються на використанні узгоджених стандартів.

Наприклад, це можуть бути зв'язки наборів даних з:

- дослідниками/ авторами,
- організацією, де виконувалось дослідження,
- грантовим проєктом, в рамках якого отримані дослідницькі результати,
- сервісами/ програмними інструментами
- та ін.

Зв'язані дані показують:

- як дані інтегруються з іншими даними
- які є контекстуальні аспекти даних
- взаємодію між сервісами та робочими процесами

Reusability— дані готові до використання в подальших дослідженнях

Повторне використання можливе лише тоді, коли зрозумілі права, контекст і якість даних.

Для цього потрібні:

- Ліцензія на використання даних
- Документація та метадані, як описують:
 - Походження даних (Data Provenance)
 - Методи збору та обробки даних
 - Умови та обмеження та використання
- Стандартизація - використання загальноприйнятих форматів, термінологій, онтологій

Інформація про походження даних (Data Provenance)

Відображає історію і методологію створення набору даних, а також контекст:

- Чому були зібрані дані
- Використані методи збору даних
- Методологія, яка була використана

- Де були зібрані дані
- Коли збирались дані
- Хто саме збирав дані
- Зв'язані дані

- Автор/ організація
- Програмне забезпечення/ інструменти
- Методи обробки даних