Управління науковими даними досліджень

Управління даними досліджень (Research Data Management, RDM) є важливою складовою організації процесу наукового дослідження. Дані дослідження можуть бути представлені у вигляді тексту, вимірювань експериментів, статистичних таблиць, мультимедіа (аудіо, відео, фото), геопросторових записів, програмного коду.

Управління даними включає планування і розробку дослідницького проєкту; збір, аналіз, зберігання даних; документування, архівування, організація доступу; повторне використання даних.

Про управління даними досліджень

Дослідницькі дані – це доказова база, що створюється або збирається дослідником для підтвердження наукових результатів. Вони формують основу аналізу досліджуваних явищ та є фундаментом для побудови всіх подальших висновків.

Дослідницькі дані можна класифікувати за різними ознаками. Зокрема, їх поділяють на ***цифрові*** **та** ***фізичні.***

За ступенем обробки дані можуть бути *первинними (необробленими)* – створеними безпосередньо під час дослідження (результати експериментів, опитувань), *вторинними* – зібраними іншими дослідниками (попередні дослідження, державна статистика), та *похідними* – отриманими в результаті обробки первинних або вторинних даних (агреговані показники, очищені або проаналізовані дані).

Джерелом дослідницьких даних може бути як власна наукова робота, так і раніше опубліковані матеріали інших авторів.

Формати представлення даних

- **кількісні:** числові виміри, статистичні показники, результати опитувань та експериментів;
- **текстові та описові:** нотатки, транскрипти інтерв'ю, матеріали фокус-груп, польові щоденники та лабораторні журнали;
- **мультимедійні:** цифрові зображення, аудіо- та відеозаписи спостережень чи процесів;
- **технічні та структурні:** математичні моделі, географічні карти, схеми, діаграми та результати 3D-моделювання;
- **програмний код та ПЗ:** алгоритми, скрипти, вихідний код додатків та спеціалізоване програмне забезпечення;
- **документація та метадані:** дослідницькі протоколи, описи методик, стандарти та інформація про структуру наборів даних.

Управління дослідницькими даними

Грантодавці, видавці та університети дедалі частіше впроваджують стандарти належного управління дослідницькими даними та їх подальшого поширення. У зв'язку з цим важливою складовою організації сучасного наукового дослідження є ефективне управління дослідницькими даними.

Управління науковими (дослідницькими) даними (Research Data Management, RDM) – це процес, що охоплює всі етапи дослідження: планування, збір, організацію, обробку, документування, зберігання, спільне використання та поширення дослідницьких даних. Метою такого управління є забезпечення ефективного використання даних, їх доступності, надійного зберігання та можливості подальшого використання.

Чому важливо керувати даними досліджень:

- **відповідність стандартам:** дотримання вимог університету, грантодавців та видавців;
- **надійність та безпека:** захист результатів від втрати та забезпечення цілісності матеріалів;
- **дотримання правових та етичних норм**: управління даними сприяє дотриманню законодавства, етичних стандартів досліджень і правил роботи з персональними або конфіденційними даними, а також сприяє захисту авторських прав;
- **оптимізація часу:** впорядкування дослідження, узгодженість процесів між виконавцями та швидкий пошук потрібної інформації;
- **видимість:** підвищення видимості результатів дослідження та зростання цитованості автора завдяки доступності та відкритості його результатів;
- **співпраця:** можливість спільної роботи з даними, їх обміну та повторного використання іншими дослідниками.

Компоненти управління дослідницькими даними

- **Планування даних** – підготовка та розробка плану управління даними (Data Management Plan, DMP), у якому визначається, які дані будуть створюватися або збиратися, як вони організовуватимуться, зберігатимуться та як ними можна буде ділитися.
- **Організація даних** – структурування файлів даних і метаданих таким чином, щоб забезпечити їхню зрозумілість, зручний доступ і можливість подальшого використання.
- **Безпека даних** – вибір відповідних організаційних рішень для захисту даних від втрати, пошкодження або несанкціонованого доступу, включаючи резервне копіювання та контроль доступу.
- **Збереження даних** – впровадження стратегій довгострокового зберігання та підтримання доступності даних, зокрема використання надійних сховищ і репозитаріїв.
- **Відповідність та етика** – забезпечення того, щоб управління даними відповідало чинному законодавству, нормативним вимогам і етичним стандартам, особливо щодо роботи з персональними або конфіденційними даними.
- **Обмін та поширення даних** – надання доступу до даних іншим дослідникам, що може включати публікацію наборів даних у відкритих репозитаріях або організацію контрольованого доступу для певних користувачів.

Життєвий цикл дослідницьких даних

Життєвий цикл даних – це послідовність взаємопов'язаних етапів, які проходять дослідницькі дані в межах дослідження: від планування, збору та обробки до довгострокового зберігання, архівації, відкритого доступу та повторного використання. Ці етапи утворюють безперервний цикл: результати завершеного дослідження після архівування стають джерелом ідей для подальшої наукової діяльності.

Життєвий цикл

Крок 1. Сплануйте роботу з даними: визначте, які дані ви будете отримувати, як їх збирати, обробляти та де зберігати. Підготуйте план управління даними (DMP).

Крок 2. Зберіть або створіть дані, дотримуючись принципів FAIR: забезпечте систематичний збір даних із одночасним створенням базових метаданих.

Крок 3. Обробіть і проаналізуйте дані: виконайте аналіз та підготуйте результати.

Крок 4. Забезпечте надійне збереження даних: використовуйте формати та сховища, що забезпечують довгострокове збереження даних.

Крок 5. Опублікуйте та поширте дані: спочатку визначте, які дані можна відкривати з урахуванням етичних і правових вимог, а потім надайте до них доступ через репозитарії або публікації.

Крок 6. Забезпечте повторне використання: оберіть відповідну ліцензію, додайте заяву про доступ до даних (Data Availability Statement), розмістіть їх у відкритому репозитарії та поширте інформацію про них на наукових платформах.

FAIR-принципи

Принципи FAIR (принципи належного управління науковими (дослідницькими) даними – міжнародновизнані принципи, що передбачають забезпечення багаторазового використання наукових (дослідницьких) даних, їх доступність, здатність до взаємодії з різними типами даних (інтероперабельність) та здійснення оперативного пошуку необхідної інформації.

Принципи FAIR запроваджені у 2016 році і є фундаментом сучасної відкритої науки.

Принципи FAIR складаються з чотирьох взаємопов'язаних компонентів:

- **Findability (знайденими)** – дані та метадані мають бути легко відшукуваними як для людей, так і для комп'ютерних систем.
- **Accessibility (доступними)** – доступ до даних має бути забезпечений на чітко визначених умовах. Навіть якщо самі дані обмежені з етичних або правових причин, метадані повинні залишатися відкритими та доступними.
- **Interoperability (сумісними)** – дані та метадані можна поєднувати з іншими даними та інструментами.
- **Reusability (багаторазовими)** – дані мають опис і ліцензію, що дозволяють їх повторне використання.

Практичне -застосування принципів FAIR

Важливо знати: FAIR ≠ Open Data

Дотримання принципів FAIR не означає, що дані обов'язково мають бути повністю відкритими.

Управління дослідницькими даними базується на принципі:

«as open as possible, as closed as necessary» – дані настільки відкриті, наскільки це можливо, і настільки закриті, наскільки це необхідно.

Дані можуть мати різні рівні доступу:

- **відкриті:** публічно доступні всім користувачам без жодних перешкод;
- **з обмеженим доступом:** призначені лише для певних груп фахівців або організацій;
- **за запитом:** надаються після офіційного звернення та отримання дозволу від власника (автора);
- **закриті:** недоступні для завантаження, проте мають відкриті метадані. Це дозволяє іншим дослідникам знайти інформацію про дослідження та зрозуміти, за яких умов доступ до даних може бути надано в майбутньому.

Навіть за умови обмеженого доступу дані можуть відповідати принципам FAIR, якщо **вони правильно описані, мають чітко визначені умови доступу та можливість коректного повторного використання**.

Інструменти для оцінювання відповідності принципам FAIR

За умови дотримання принципів FAIR доцільно використовувати спеціальні інструменти для оцінювання відповідності наборів даних. Такі сервіси дозволяють перевірити відповідність даних вимогам FAIR, виявити "слабкі місця" та вдосконалити опис і умови доступу до даних перед поданням статті до журналу або заявки на грантовий проєкт.

FAIR-Aware (https://fairaware.dans.knaw.nl/) – онлайн-інструмент самооцінювання у вигляді інтерактивного опитувальника, який допомагає дослідникам зрозуміти вимоги FAIR ще до завантаження даних у сховище. Він фокусується на підвищенні обізнаності та надає практичні поради для кожного кроку управління дослідницькими даними.

Ресурс можна використовувати на будь-якому етапі дослідження, зокрема під час планування або перед розміщенням даних у репозитарії. Опитування складається з 10 запитань і триває від 10 до 30 хвилин, після чого користувач отримує персоналізовані рекомендації щодо вдосконалення практик FAIR.

FAIR Data Self-Assessment Tool **(**https://ardc.edu.au/resource/fair-data-self-assessment-tool/**)** – зручний інструмент для швидкої перевірки відповідності конкретного набору дослідницьких даних.

Інструмент доцільно використовувати перед публікацією або розміщенням даних у репозиторії. Користувач відповідає на серію запитань щодо пошуку, доступності, сумісності та повторного використання набору даних, після чого сервіс підбиває підсумок і вказує на прогалини, які варто виправити.

F-UJI **(**https://www.f-uji.net/?action=test**)** – автоматизований інструмент для оцінювання відповідності наборів дослідницьких даних принципам FAIR на основі метаданих.

Ресурс призначений для аналізу вже опублікованих наборів даних. Достатньо ввести ідентифікатор (наприклад, DOI), і F-UJI представить детальний звіт з оцінкою кожного принципу та рекомендаціями щодо покращення технічних параметрів. Розроблений згідно з вимогами **EOSC** (European Open Science Cloud) (https://eosc.eu/).

EUDAT FAIR Data Checklist (https://zenodo.org/records/1065991**)** – практичний чекліст для самоперевірки на різних етапах роботи з даними.

Допомагає покроково контролювати виконання вимог FAIR під час планування та підготовки до публікації. Розроблений у межах ініціативи **EUDAT** (European e-Infrastructure Data Infrastructure) (https://eudat.eu/?utm_source=chatgpt.com) та доступний через Zenodo.

План управління даними

План управління даними (Data Management Plan, DMP) – це документ, який описує, як дослідник або група дослідників будуть обробляти, організовувати, зберігати та поширювати наукові (дослідницькі) дані протягом усього дослідження та після його завершення.

DMP – це не формальність, а стратегія роботи з даними протягом усього життєвого циклу дослідження.

Коли потрібен DMP

DMP є обов’язковим або рекомендованим у таких випадках:

- участь у міжнародних та національних грантових проєктах;
- проведення досліджень із персональними або чутливими даними;
- реалізація політики відкритої науки в університеті.

Що має бути в DMP

План управління даними зазвичай включає (структура може відрізнятися залежно від вимог грантодавця):

1. Опис даних:

- Типи даних (текстові, числові, аудіо, відео тощо)
- Формати файлів
- Обсяг даних
- Джерела отримання

2. Документація та метадані:

- Стандарти метаданих
- Способи опису даних
- Версіонування (фіксація, відстеження та управління змінами)

3. Зберігання та резервне копіювання:

- Де зберігатимуться дані
- Як здійснюється резервне копіювання
- Хто має доступ

4. Правові та етичні аспекти:

- Персональні дані
- GDPR (General Data Protection Regulation) — загальний регламент ЄС про захист персональних даних
- Інформована згода
- Ліцензування

5. Доступ та повторне використання (відповідно до FAIR-принципів):

- Чи будуть дані відкритими
- Де вони будуть опубліковані
- Умови повторного використання
- Репозитарій для розміщення

6. Архівування та довгострокове зберігання:

- Термін зберігання
- Формати для архівування
- Відповідальна особа

Чому важливий DMP

- підвищує якість дослідження;
- допомагає відповідати вимогам донорів;
- захищає персональні дані;
- полегшує повторне використання результатів;
- сприяє впровадженню політики відкритої науки.

Як створити план управління даними

Для створення плану управління даними можна скористатися такими інструментами: **DMPonline, DMPTool, ARGOS.**

Основні можливості DMPonline:

- розроблений Digital Curation Centre (Велика Британія);
- має готові шаблони для різних грантодавців;
- дозволяє експортувати ваш план у форматі Word, Excel або PDF;
- підтримує командну роботу.

Основні можливості DMPTool:

- створення плану управління даними;
- шаблони від фінансових установ та організацій;
- командна робота над документом;
- експорт документів;
- збереження та архівування планів.

Основні можливості ARGOS:

- інструмент від OpenAIRE;
- інтегрований з європейською інфраструктурою відкритої науки;
- підтримує FAIR-принципи;
- підходить для проєктів Horizon Europe;

Організація та опис даних

Організація та збереження даних

Організація даних – ефективна робота з даними передбачає структуризацію файлів, документування процесу дослідження та метаданих.

Структуризація файлів потребує створення картотеки проєкту та окремих папок з джерелами публікації, статистичними (експериментальними) даними, результатами досліджень (програмний код, таблиці аналізу даних, текст роботи) та README файлом (текстовий файл, який містить основну інформацію про проєкт, програму або архів, супроводжуючи їх при розповсюдженні).

Документування метаданих, методології аналізу і трансформації даних забезпечує розуміння даних та процесу дослідження усіма зацікавленими сторонами.

Метадані – інформація про оригінальні дані, що описують та допомагають класифікувати, упорядковувати та характеризувати дані. Ключовими елементами метаданих є визначення та позначення показників, одиниць їх виміру, короткий опис методології оцінювання та джерел даних.

Назви файлів мають бути унікальні, змістовні, не дуже довгі. Бажано використовувати стандартизовану форму для різних версій документів.
Рекомендовані елементи для назви файлів:

➡ назва проєкту або ім'я дослідника

➡ вид роботи або дата створення файлу (YYYYMMDD)

➡ версія документа (напр., V1, V1_2, V2)

➡ використання символів з наборів A-Z, a-z, 0-9, дефіс, підкреслення і крапка

Приклади: MultivariteAnalysis_Part2_20190221.docx, Protsiuk_Thesis_V1.pdf, UkrStat _2000-2019.xlsx

Для забезпечення машинного читання файлів використовують такі **формати даних**:

➡ Табличні дані – CSV замість XLSX

➡ Текстові дані – TXT або PDF замість DOC

➡ Бази даних – XML або SQLITE замість MDB, DBF, SQL

➡ Візуальні – PDF, TIFF, JPEG2000, MPEG-4, WAVE, AIFF

Збереження даних

Для збереження даних та їх відтворення на випадок пошкодження використовують резервне копіювання інформації.

Правила резервного копіювання 3-2-1:

➡ 3 копії (1 оригінал READ ONLY, 2 копії)

➡ 2 різні типи сховищ (жорсткий диск, USB, хмара)

➡ 1 копія на фізичному носії, 1 копія на е-диску

Спеціальні програми для управління проєктами та версіями файлів: GIT: GitHub, GitLab, BitBucket, Trello.

Платформи для зберігання та обміну файлами: Open Science Framework, Google Drive, Dropbox, Box.

Зберігання даних

Де і як зберігати дані

Належне зберігання даних досліджень є важливою складовою їх ефективного управління та запобігає несанкціонованому доступу, небажаним змінам даних, їх розголошенню або знищенню протягом усього життєвого циклу наукового проєкту. Вибір місця та способу зберігання залежить від типу даних, рівня їх чутливості, обсягу та вимог установи або грантодавця.

Де зберігаються дані наукових досліджень

1. Локальні носії інформації (на флешці, HDD- чи SSD-дисках тощо).

Переваги: зручність для оперативної роботи; доступ до даних є навіть без інтернету.

Недоліки: доступ до даних тільки при наявності носія; носій може загубитися, зламатися або застаріти; обсяг даних обмежений можливостями носія.

2. Захищені сервери установ (репозитарії, цифрові бібліотеки, хмарні архіви даних)

Переваги: централізоване зберігання; захист від втрати даних при проблемах з офісною технікою; власник серверу повністю контролює все, що відбувається з файлами; зручний доступ для спільного використання даних як у межах установи, так і поза нею.

Недоліки: залежність від інтернету; втрата даних при фізичному знищенні техніки.

3. Сертифіковані хмарні сервіси (Amazon S3, Google Диск, OpenScienceFramework, Dropbox і ін.)

Переваги: доступ до даних з будь-якого пристрою за умови наявності інтернету; автоматичне резервне копіювання даних; забезпечення безпеки даних, включаючи шифрування, автентифікацію та інші заходи; зручні при співпраці з партнерами з інших організацій.

Недоліки: залежність від інтернету; повна залежність від політики безпеки та працездатності провайдера; ризики виникнення інцидентів безпеки, таких як витоки даних або атаки на систему.

Захист даних та резервні копії

Захист даних дослідження є ключовою складовою належного управління науковою інформацією та передбачає забезпечення їхньої конфіденційності, цілісності й доступності протягом усього життєвого циклу дослідження. Надійний захист особливо важливий у випадках роботи з персональними, чутливими або конфіденційними даними.

Захист даних дослідження – це комплекс правових, організаційних та технічних заходів, спрямованих на забезпечення конфіденційності, цілісності та доступності інформації протягом усього її життєвого циклу. Він запобігає несанкціонованому доступу, зміні, розкриттю чи знищенню даних, що є критичним для наукової етики та безпеки.

Важливе значення при захисті даних має **резервне копіювання або бекап**.

Резервне копіювання (бекап) – це створення копій інформації для відновлення у разі втрати. Основні принципи включають використання різних носіїв (хмари, зовнішнього диска) та автоматизацію резервного копіювання.

Стратегії резервного копіювання

Правильна організація резервного копіювання вимагає ретельного планування та впровадження ефективних стратегій, що дозволяють зберегти дані в умовах різноманітних загроз. Однією з таких стратегій є **правило 3-2-1:**

- *3 копії* (1 оригінал READONLY, 2 копії);
- *2 різні типи сховищ* (жорсткий диск, USB, хмара);
- *1 копія поза основною локацією*, аби уникнути ризику, пов'язаного з однією фізичною точкою відмови (1 копія на фізичному носії, 1 копія на е-диску).

Однак у сучасних умовах цей підхід дедалі частіше виявляється недостатнім. Зловмисники навчилися не лише шифрувати продуктивні системи, а й цілеспрямовано знищувати або шифрувати резервні копії. Саме це стало підґрунтям для появи **модифікованого правила 3-2-1-1-0** шляхом доповнення правила 3-2-1 двома критично важливими вимогами:

- наявністю щонайменше *однієї незмінної* (immutable) або фізично відключеної копії;
- обов'язковою перевіркою відновлюваності резервних копій до стану *«нуль помилок».*

Застосування цього підходу допомагає зберегти цілісність резервних копій і забезпечити їх доступність у випадку непередбачуваних ситуацій, а також забезпечує перехід від формального резервного копіювання до повноцінної стратегії кіберстійкості.

Контроль доступу до даних

Контроль доступу до даних є важливим елементом безпеки даних, який визначає, хто має доступ до даних досліджень та захищає, таким чином, інформацію від несанкціонованого доступу, зміни або знищення.

Для управління доступом застосовуються:

- персоналізовані облікові записи користувачів;
- політика складних паролів і багатофакторна автентифікація;
- розмежування прав доступу за ролями;
- регулярний перегляд і актуалізація прав доступу.

Розподіл доступу

Робота над даними досліджень у команді передбачає чітке визначення повноважень кожного члена команди для роботи з різними об'єктами бази даних (база даних повністю, окремі таблиці, записи або значення даних). Повноваження окремих користувачів при роботі з одним і тим самим об'єктом можуть бути різними. Користувачі можуть бути об'єднані в спеціальні групи користувачів з певним рівнем доступу. Один користувач може входити до декількох груп.

Розподіл ролей доступу до файлів:

- тільки читання;
- можливість змінення;
- надання дозволів.

 **Особливу увагу слід приділяти доступу до чутливих або персональних даних.** У таких випадках доступ може бути додатково обмежений або надаватися лише після проходження відповідних процедур погодження. Також доцільно використовувати *процес фіксації подій (логування),* що дозволяє відстежувати, хто і коли працював із даними.

У разі зміни посадових обов'язків або припинення роботи з дослідженнями доступ до даних своєчасно обмежується або анулюється.

Безпечне видалення даних

Безпечне видалення даних – це процес, при якому файли видаляються так, що їх неможливо відновити навіть за допомогою спеціальних програм. Звичайне видалення не стирає інформацію з носія, а лише позначає її як видалену, що дозволяє зловмисникам відновити ці дані.

Безпечне видалення даних дослідження є важливим етапом управління дослідницькою інформацією, що забезпечує захист конфіденційності, дотримання етичних норм та вимог законодавства. Після завершення проєкту або закінчення терміну зберігання даних виникає необхідність їх належного знищення, щоб унеможливити несанкціонований доступ або відновлення.

Принципи видалення даних

1. *Перед видаленням даних* слід переконатися, що вони більше не мають наукової, юридичної чи архівної цінності. Також необхідно перевірити, чи не підпадають ці дані під політики відкритого доступу або вимоги щодо довготривалого зберігання.

Безпечне видалення передбачає використання **спеціальних методів, які унеможливлюють відновлення інформації.** До них належать:

- перезапис даних (overwrite) із застосуванням спеціалізованого програмного забезпечення;
- криптографічне стирання (знищення ключів шифрування);
- фізичне знищення носіїв (подрібнення, спалення, демагнітизація для магнітних носіїв).

2. Необхідно подбати про те, щоб *усі копії* відповідного набору даних (включаючи будь-які резервні копії) були ідентифіковані та оброблені належним чином.

3. Особливу увагу слід приділяти *персональним даним учасників дослідження.* Їх видалення має здійснюватися відповідно до принципів конфіденційності та з урахуванням вимог законодавства про захист персональних даних.

4. Доцільно *документувати процес видалення даних:* фіксувати дату, спосіб і відповідальних осіб. Це забезпечує прозорість та можливість аудиту дослідницької діяльності.

Репозитарій для зберігання даних

https://repository.ldufk.edu.ua/home

Обмін даними та повторне використання

Обмін дослідницькими даними – це не лише виконання формальних вимог, а й спосіб зробити наукове дослідження більш помітним і вагомим. Відкритість забезпечує прозорість науки та підвищує довіру до наукових результатів.

Поширення дослідницьких даних здійснюється за принципом ***«as open as possible, as closed as necessary»*** **з дотриманням принципів FAIR.**

Ключові переваги поширення дослідницьких даних

- **Повторне використання та розвиток науки:** відкриті дані дають змогу використовувати результати досліджень у нових наукових роботах, поєднувати їх з іншими наборами даних та створювати нові знання.
- **Підвищення видимості та цитованості:** розміщення даних у відкритих репозитаріях підвищує видимість дослідження, сприяє зростанню цитованості та забезпечує присвоєння постійних ідентифікаторів (зокрема DOI).
- **Відповідність національним та міжнародним вимогам:** відкриття даних відповідає політиці відкритої науки в Україні (зокрема вимогам щодо FAIR-даних для досліджень, що фінансуються з державного бюджету) та вимогам міжнародних грантодавців, таких як Horizon Europe, які передбачають обов'язковий обмін даними.
- **Прозорість і довіра:** доступність даних забезпечує можливість перевірки та відтворення результатів досліджень, що підвищує довіру до наукових результатів.
- **Підвищення якості досліджень:** вимоги наукових журналів щодо відкритості первинних (сирих) даних сприяють підвищенню якості досліджень завдяки можливості повторного аналізу результатів та виявлення помилок.
- **Надійність і збереження даних:** публічні репозитарії забезпечують довгострокове зберігання, резервне копіювання та вищий рівень захисту даних порівняно з локальними носіями.

Що враховувати при поширення дослідницьких даних

- **Конфіденційність та етика:** забезпечити захист персональних і чутливих даних (зокрема шляхом анонімізації) та дотримання умов інформованої згоди учасників дослідження.
- **Права інтелектуальної власності:** переконатися у наявності прав на використання та поширення даних, а також врахувати обмеження, пов'язані з умовами договорів і фінансування.
- **Ліцензування:** обрати відповідну ліцензію (наприклад, Creative Commons), яка визначає умови доступу та повторного використання даних іншими користувачами. У разі використання вторинних даних необхідно перевірити умови їх ліцензування.

- **Якість даних:** поширювати дані, що представлені у відкритих або широко використовуваних у відповідній галузі форматах, повні, точно задокументовані та зрозумілі як для людини, так і для машинної обробки.

Способи поширення дослідницьких даних

- Публікація даних як додаткових матеріалів до наукової статті.
- Розміщення у відкритих мультидисциплінарних або галузевих репозитаріях (наприклад, **DataverseUA, Zenodo, Figshare, Dryad**).
- Публікація у наукових журналах даних (наприклад, **Scientific Data** від Springer Nature, **Data in Brief** від Elsevier).

У випадках, коли дані не можуть бути оприлюднені у відкритому доступі через етичні, правові або договірні обмеження, вони можуть поширюватися через репозитарії з контрольованим доступом або залишатися закритими. При цьому у відкритому доступі мають залишатися метадані, що описують ці дані.

Конфіденційні та персональні дані: правила публікації

Не всі дані, отримані в межах дослідження, обов'язково мають бути оприлюднені. Вибір рівня доступу до дослідницьких даних має враховувати етичні вимоги, законодавство про захист персональних даних, умови договорів та політику університету. Публікація окремих типів даних може порушувати права учасників дослідження або суперечити етичним нормам і законодавчим вимогам.

Не підлягають оприлюдненню дані, що потребують захисту з юридичних або етичних причин. До них належать, зокрема, конфіденційні персональні дані, псевдонімізовані дані, а також інші види інформації з обмеженим доступом.

Під час відкриття дослідницьких даних необхідно враховувати різні типи конфіденційних та чутливих даних, які не підлягають відкритому поширенню без спеціальних правових або етичних підстав.

Категорії конфіденційних та чутливих даних

- **персональні дані**, за допомогою яких можна прямо або опосередковано ідентифікувати особу (імена, адреси, контактні дані, ідентифікаційні номери);
- **дані спеціальної категорії** (про здоров'я, генетичні та біометричні дані, расове або етнічне походження, політичні, релігійні чи філософські переконання, сексуальну орієнтацію);
- **дані про кримінальні правопорушення** та пов'язані з ними судові або безпекові відомості;
- **культурно чутливі дані**, публікація яких може завдати шкоди окремим особам або спільнотам;
- **екологічно чутливі дані**, зокрема інформація про місцезнаходження рідкісних або зникаючих видів;
- **комерційно конфіденційні дані**, що мають обмеження через договори, фінансування або інтелектуальну власність;
- **дані з етичними, правовими або безпековими обмеженнями.**

Підготовка до публікації дослідницьких даних

Перед публікацією дослідницьких даних доцільно застосовувати **процедури анонімізації або псевдонімізації,** які зменшують ризик ідентифікації учасників дослідження та дозволяють безпечно поширювати дані.

Якщо доступ до даних обмежений, у репозитарії мають бути доступні їхні метадані з детальним описом умов отримання даних (наприклад, на період ембарго або за запитом). Такий підхід дає змогу дотримуватися принципів FAIR, не порушуючи етичних і правових норм.

Водночас доцільно розглянути обмін такими даними, які:

- підкріплюють наукові публікації та підтверджують отримані результати;
- є унікальними або складними для відтворення (польові спостереження, опитування, експериментальні вимірювання);
- можуть бути корисними для інших дослідників;
- не дозволяють ідентифікувати конкретних осіб (анонімізовані або узагальнені);
- не містять правових, етичних або договірних обмежень для поширення.

Перед публікацією даних також важливо перевірити вимоги грантодавця, оскільки окремі програми фінансування встановлюють обов'язкові правила щодо обміну дослідницькими даними.

Інформована згода учасників дослідження

Інформована згода є ключовою етичною вимогою для досліджень, які проводяться за участю людей. Вона має враховуватися протягом усього життєвого циклу дослідження – від етапу планування до публікації, архівування та поширення даних. Інформована згода означає, що учасники дослідження свідомо та добровільно погоджуються на участь, а також розуміють, як саме їхні дані будуть використовуватися.

Згода на участь у дослідженні

Згода на участь має бути отримана до початку проведення дослідження. Вона передбачає надання учасникам зрозумілої та повної інформації про:

- мету дослідження;
- характер і обсяг участі;
- можливі переваги та потенційні ризики;
- право відмовитися від участі у будь-який момент без негативних наслідків;
- умови конфіденційності та захисту приватного життя.

Згода на використання на поширення даних

Вважливо інформувати учасників про те, як саме будуть використовуватися зібрані дані, як у межах дослідження, так і після його завершення. Це включає інформацію про:

- типи даних, що збираються, умови та строки їх зберігання;
- способи використання даних (публікація, поширення та повторне використання);
- форми поширення даних (анонімізовані, стенограми, аудіо- чи відеоматеріали тощо);
- коло осіб або організацій, які можуть отримати доступ до даних, а також контактну інформацію.

Інформована згода повинна бути належним чином задокументована. Зазвичай вона оформлюється у вигляді **інформаційного листа та форми згоди, підписаної учасником дослідження.**

Документування інформованої згоди

- учасник ознайомився з інформацією про дослідження та зрозумів її;
- йому було надано можливість поставити запитання;
- участь у дослідженні є добровільною;
- учасник має право відмовитися від участі у будь-який момент без пояснення причин і негативних наслідків;
- умови захисту конфіденційності, зокрема анонімізація або використання псевдонімів;
- перелік даних, які можуть бути використані в публікаціях (наприклад, цитати, аудіо- чи відеоматеріали);
- підписи та дати підпису учасника та дослідника.

Наявність належно оформленої інформованої згоди забезпечує правомірну публікацію, архівування та обмін дослідницькими даними.

Проведення дослідження без отримання згоди

Проведення дослідження **можливе без отримання** згоди суб'єкта даних, **якщо:**

- дослідження проводиться у громадських місцях, де відсутні очікування щодо конфіденційності;
- використовується загальнодоступна інформація про осіб (зокрема аналіз публічних записів або архівних матеріалів);
- здійснюється використання інформації з відкритих джерел (інтернет-сайтів, соціальних мереж), за умови відкритого доступу до відповідних сторінок;
- використовуються вторинні та/або анонімізовані персональні дані.

Правові та етичні вимоги

Дослідження, що передбачають участь людей, використання їхніх персональних даних або біологічних зразків, повинні отримати попередній висновок **етичної комісії** університету до початку збору даних.

Якщо учасники не надали згоди на відкритий доступ до даних, рекомендується оприлюднювати метадані та умови доступу до таких даних.

В Україні інформована згода учасників дослідження є не лише етичною вимогою, а й правовою умовою для обробки персональних даних відповідно до Закону України «Про захист персональних даних» та вимог Загального регламенту Європейського Союзу про захист даних (GDPR Regulation (EU) 2016/679).

Учасники дослідження мають право:

- отримувати інформацію про обробку своїх персональних даних;
- отримувати доступ до своїх персональних даних;
- відкликати надану згоду на обробку даних;
- вимагати виправлення або видалення (право на забуття) своїх даних у випадках, передбачених законодавством.

Обробка біоматеріалів від людей-учасників має здійснюватися відповідно до чинного Порядку збору, зберігання та використання біологічних зразків людини з дослідницькою метою, затвердженого Кабінетом Міністрів України.

Анонімізація даних та контроль доступу

Анонімізація – один із основних методів захисту конфіденційності даних. Це процес обробки даних, спрямований на видалення або зміну інформації, яка дозволяє прямо чи опосередковано ідентифікувати учасників дослідження.

Після здійснення анонімізації встановити особу учасника стає неможливо.

Анонімізація дозволяє поширювати дослідницькі дані етично та законно, забезпечуючи захист конфіденційності учасників.

Прямі та непрямі ідентифікатори

Ідентифікувати особу можуть:

Прямі ідентифікатори – інформація, що безпосередньо вказує на конкретну особу, наприклад:

- прізвище, ім'я та по батькові;
- адреса або контактні дані;
- персональні ідентифікаційні номери (паспорт, ІПН);
- аудіо- або відеозаписи та фотографії, які дозволяють ідентифікувати особу.

Непрямі ідентифікатори – дані, які можуть ідентифікувати особу у поєднанні з іншою інформацією:

- вік або дата народження;
- місце роботи, посада чи професія;
- рівень доходу та інші соціально-демографічні характеристики;
- рідкісні захворювання;
- географічні мітки (місце проживання або невеликі населені пункти).

Метод анонімізації

- видалення ідентифікаційних даних;
- заміна імен псевдонімами або кодами;
- агрегування або узагальнення даних;
- редагування текстових матеріалів (наприклад, інтерв'ю);
- обробка зображень або відео (наприклад, розмиття облич).

Анонімізацію доцільно планувати ще на етапі підготовки дослідження. Під час підготовки даних до поширення необхідно оцінювати ризик **повторної ідентифікації.**

Поряд з анонімізацією застосовується також **псевдонімізація.**

Псевдонімізовані дані – це дані, у яких ідентифікаційна інформація замінена штучними ідентифікаторами (кодами або псевдонімами). За наявності

додаткової інформації такі дані можуть бути повторно пов’язані з конкретною особою.

Інструменти для анонімізації даних

Для підготовки дослідницьких даних до публікації або поширення можна використовувати спеціальні інструменти та шаблони:

- **ARX Data Anonymization Tool** – програмне забезпечення для анонімізації табличних даних та оцінювання ризику повторної ідентифікації.
- **Amnesia** – веб-інструмент для анонімізації наборів даних, розроблений у межах інфраструктури OpenAIRE.
- **Anonymisation Plan Template** – шаблон плану анонімізації, розроблений Finnish Social Science Data Archive (FSD).

Детальні інструкції щодо ефективної анонімізації дослідницьких даних наведені у матеріалах провідних організацій у сфері управління дослідницькими даними:

- **UK Data Service – Anonymisation**
- **Information Commissioner's Office (ICO) – Anonymisation guidance**

Ліцензування даних

Ліцензування є важливим етапом поширення дослідницьких даних, оскільки визначає умови доступу, використання та повторного використання даних іншими користувачами.

Дослідницькі дані, а також пов’язані з ними матеріали (тексти, зображення, програмний код, бази даних) можуть бути об’єктами авторського права або суміжних прав. Авторське право на такі матеріали виникає автоматично з моменту їх створення і не потребує спеціальної реєстрації. За замовчуванням діє принцип «усі права застережено», що означає: інші особи не можуть використовувати і поширювати ці матеріали без дозволу автора.

Щоби забезпечити законне повторне використання дослідницьких даних, автори можуть застосовувати відкриті ліцензії.

Ліцензії Creative Commons

Найпоширенішими відкритими ліцензіями для дослідницьких даних є **ліцензії Creative Commons,** які дозволяють авторам визначати умови доступу, використання та поширення їхніх матеріалів.

Існує шість основних типів ліцензій Creative Commons. Усі вони передбачають **обов’язкове зазначення авторства,** але можуть містити додаткові обмеження щодо використання матеріалів.

Назва ліцензії	Умови використання
CC BY	дозволяє використовувати, поширювати та змінювати твір або дані досліджень за умови обов'язкового зазначення автора або творця даних
CC BY-SA	дозволяє поширювати або змінювати оригінальний твір або дані досліджень із зазначенням автора та за умови поширення похідних матеріалів на тих самих умовах
CC BY-NC	дозволяє використовувати, поширювати та змінювати твір або дані досліджень лише в некомерційних цілях із зазначенням автора
CC BY-ND	дозволяє копіювати та поширювати твір або дані досліджень без будь-яких змін із зазначенням автора; комерційне використання дозволено
CC BY-NC-SA	дозволяє повторно поширювати та використовувати твір або дані досліджень в некомерційних цілях із зазначенням автора та на умовах тієї ж ліцензії
CC BY-NC-ND	дозволяє лише некомерційне використання без внесення будь-яких змін із обов'язковим зазначенням автора

Крім того, існує **ліцензія CC0,** яка дозволяє передати твір або дані досліджень у публічне надбання та використовувати їх без обов'язкового зазначення автора.

Правильний вибір ліцензії дозволяє не лише захистити права автора, а й забезпечити відкритість і доступність дослідницьких даних, що відіграє важливу роль для розвитку відкритої науки. **Як обрати машиночитну ліцензію Creative Commons?**

Ідентифікатор DOI

Для забезпечення надійної ідентифікації та доступності дослідницьких даних важливо використовувати постійні ідентифікатори, зокрема **DOI (Digital Object Identifier).**

DOI – це унікальний цифровий ідентифікатор, який присвоюється науковим публікаціям, наборам даних та іншим результатам досліджень. Він забезпечує стабільне посилання на набір даних незалежно від змін його місцезнаходження в мережі.

Переваги використання DOI

- забезпечує довготривалий доступ до дослідницьких даних
- дає змогу коректно цитувати результати досліджень
- підвищує видимість і впізнаваність наукових матеріалів
- дозволяє відстежувати використання та цитування даних

 Наявність DOI є важливою умовою для публікації дослідницьких даних у відкритих репозитаріях і сприяє їх інтеграції в міжнародний науковий простір.

Джерело: https://mon.gov.ua/static-objects/mon/sites/1/nauka/2024/12/31/metod-rekomendatsiyi-shchodo-upravlinnya-naukovymy-danymy-31-12-2024.pdf