НАЦІОНАЛЬНА АКАДЕМІЯ ПЕДАГОГІЧНИХ НАУК УКРАЇНИ

ІНСТИТУТ ЦИФРОВІЗАЦІЇ ОСВІТИ

ФОРМУВАННЯ ТА ВИКОРИСТАННЯ FAIR-ДАНИХ У ГАЛУЗІ ОСВІТНІХ НАУК

Довідник



КИЇВ 2026

УДК 37.01:001.891-021.465-047.44:004
Ф79

Рекомендовано до друку вченою радою
Інституту цифровізації освіти НАПН України
(протокол № 14 від 29.10.2026 р.)

Рецензенти:

Кузьмінська О. Г.	доктор педагогічних наук, професор, професор кафедри інформаційних систем і технологій, Національний університет біоресурсів і природокористування України
Гаманюк В. А.	доктор педагогічних наук, професор, проректор з наукової роботи, професор кафедри німецької мови і літератури з методикою викладання, Криворізький державний педагогічний університет

Ф79 Формування та використання FAIR-даних у галузі освітніх наук : довідник / Спірін О. М., Іванова С. М., Вакалюк Т. А., Вознюк Н. В., Кільченко А. В., Мінтій І. С., Новицька Т. Л., Олексюк В. П., Осадча К. П., Шиненко М. А. [за ред. О. М. Спіріна, В. П. Олексюка]. Київ : ІЦО НАПН України, 2026. 110 с.

ISBN 978-617-8330-60-6 – PDF

Довідник присвячено теоретичним і практичним засадам формування та використання FAIR-даних у галузі освітніх наук. У ньому розглянуто принципи FAIR – відшукуваність, доступність, сумісність і повторне використання даних, типи освітніх наукових даних, їх життєвий цикл та особливості управління ними. Висвітлено питання стандартизації FAIR-даних, розроблення планів управління даними, вибору форматів і цифрових сервісів для зберігання, доступу, ліцензування та публікації даних. Окрему увагу приділено захисту персональних даних, етичним і правовим аспектам, використанню репозитаріїв та можливостям Європейської хмари відкритої науки. Довідник містить практичні засади пошуку, доступу, інтеграції, аналізу та повторного використання FAIR-даних у педагогічних дослідженнях, зокрема із застосуванням сучасних цифрових технологій та інструментів освітньої аналітики. Наведено приклади використання FAIR-даних в опитуваннях, психолого-педагогічному оцінюванні, навчальній аналітиці та інших видах досліджень.

Довідник призначено для наукових і науково-педагогічних працівників, аспірантів, докторантів, здобувачів освіти, фахівців наукових установ і закладів вищої освіти та всіх, хто займається організацією, зберіганням, публікацією, аналізом і повторним використанням дослідницьких даних у галузі освітніх наук.

УДК 37.01:001.891-021.465-047.44:004

ISBN 978-617-8330-60-6 – PDF

© Спірін О. М., Іванова С. М., Вакалюк Т. А., Вознюк Н. В., Кільченко А. В., Мінтій І. С., Новицька Т. Л., Олексюк В. П., Осадча К. П., Шиненко М. А. © ІЦО НАПН України, 2026

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ 5

ПЕРЕДМОВА 6

Р О З Д І Л І ФОРМУВАННЯ FAIR-ДАНИХ У ГАЛУЗІ ОСВІТНІХ НАУК 9

1.1. Принципи FAIR: концептуальні засади 9

1.1.1. Виникнення та еволюція концепції FAIR 9
1.1.2. Принцип Findable (відшукуваність) 12
1.1.3. Принцип Accessible (доступність) 12
1.1.4. Принцип Interoperable (сумісність) 14
1.1.5. Принцип Reusable (придатність до повторного використання) 17

1.2. Типи освітніх наукових даних та їх класифікація 18

1.2.1. Кількісні дані 18
1.2.2. Якісні дані 19
1.2.3. Змішані типи даних 20
1.2.4. Адміністративні та інституційні освітні дані 21
1.2.5. Класифікація за типами і методами дослідження 22

1.3. Життєвий цикл даних і суб'єкти управління FAIR-даними 23

1.3.1. Етапи життєвого циклу дослідницьких даних 23
1.3.2. Розподіл ролей і відповідальності в управлінні дослідницькими даними 24

1.4. Стандартизація FAIR-даних в освітніх науках 26

1.4.1. Міжнародні стандарти управління науковими даними 26
1.4.2. Освітні метадані та стандарти 27
1.4.3. Модель стандартизації FAIR-даних для галузі освітніх наук: рівні, критерії та показники оцінювання 28
1.4.4. Зіставлення з міжнародними інструментами оцінювання (RDA FAIR Maturity Model, FAIRsFAIR, F-UJI) 33

1.5. Плани управління даними (Data Management Plans, DMP) 34

1.5.1. Структура і вимоги DMP 34
1.5.2. Шаблон DMP для педагогічного дослідження (заповнювана форма) 35

1.6. Формати даних та їх зберігання 38

1.6.1. Відкриті й пропрієтарні формати за типами освітніх даних 39
1.6.2. Порівняльна характеристика форматів за придатністю до довготривалого зберігання 40

1.7. Цифрові сервіси та репозитарії для зберігання FAIR-даних 41

1.7.1. Загальнонаукові репозитарії даних 42
1.7.2. Спеціалізовані освітні сховища та відкриті бази даних 44
1.7.3. Платформи інституційних репозитаріїв і бібліотечних систем: порівняння за критеріями FAIR 44
1.7.4. Національна та інституційна інфраструктура 45
1.7.5. Захищені інституційні архіви для чутливих освітніх даних: досвід FAIRVault 46
1.7.6. Європейська хмара відкритої науки (EOSC) та її сервіси для українських дослідників 47

1.8. Доступ, ліцензування та права 48

1.8.1. Ліцензування: Creative Commons та Open Data Commons (ODC) 48
1.8.2. Автентифікація й авторизація доступу 50

1.9. Етичні та правові аспекти захисту персональних даних 50

1.9.1. Нормативна база: GDPR і Закон України «Про захист персональних даних» 50

1.9.2. Анонімізація та псевдонімізація: методи й інструменти (ARX, Amnesia, sdcMicro, OpenDP) 51
1.10. Пакування, ідентифікація та публікація даних 53
1.10.1. Стандарти пакування дослідницьких об'єктів (RO-Crate, BagIt) 54
1.10.2. Присвоєння постійних ідентифікаторів (DOI, ORCID, ROR) 54
1.10.3. Підготовка заяви про доступність даних (Data Availability Statement) 55
1.10.4. Завантаження та верифікація даних у репозитарії 56

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ДО РОЗДІЛУ 1 57

Р О З Д І Л І І ВИКОРИСТАННЯ FAIR-ДАНИХ У ГАЛУЗІ ОСВІТНІХ НАУК 64

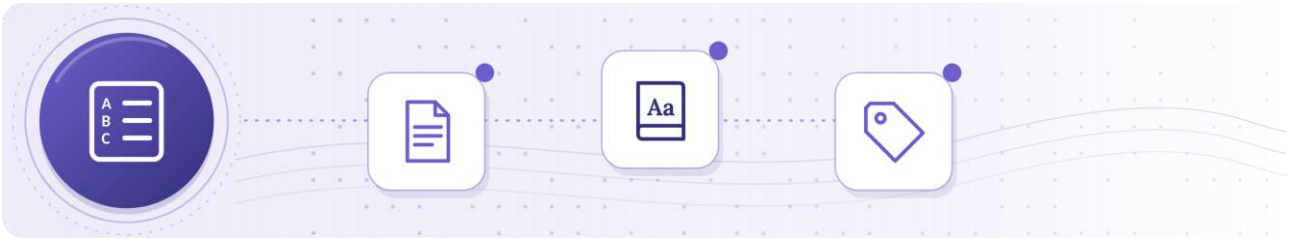
2.1. Пошук та доступ до FAIR-даних 65
2.1.1. Стратегії пошуку наукових даних (Google Dataset Search, DataCite Search, BASE, OpenAIRE) 65
2.1.2. Оцінювання якості, прав та приватності знайдених даних 66
2.1.3. Правові та етичні аспекти доступу 67
2.2. Технічна і семантична інтеграція FAIR-даних у дослідженні 68
2.2.1. Узгодження форматів і термінології даних з різних джерел 68
2.2.2. Інтеграція зовнішніх даних у власне дослідження 69
2.2.3. Цитування, походження даних і відтворюваність дослідження 70
2.3. Аналіз, вторинне використання та метадослідження на основі FAIR-даних 71
2.3.1. Вторинний аналіз освітніх даних, зокрема методами машинного навчання 71
2.3.2. Методи кількісного й якісного аналізу; інструменти освітньої аналітики 72
2.3.3. Метааналіз, систематичні огляди та міжнародні порівняльні дослідження 74
2.3.4. Розроблення доказової освітньої політики на основі FAIR-даних 75
2.4. Приклади застосування FAIR-даних у педагогічних дослідженнях 77
2.4.1. Опитування та психолого-педагогічне оцінювання 77
2.4.2. Навчальна аналітика і захист персональних даних 79
2.4.3. Синтетичні дані як інструмент етичного повторного використання 80
2.4.4. Покроковий приклад створення FAIR-пакета 81
2.5. Цифрова компетентність дослідника та інституційне впровадження 84
2.5.1. Складники цифрової компетентності наукового працівника 84
2.5.2. Відповідальне управління даними: етика та академічна доброчесність 85
2.5.3. Підготовка нових дослідників: навчання на реальних масивах даних 87
2.5.4. Ресурси й освітні платформи (The Turing Way, FOSTER+, OpenAIRE, Рамка компетентностей FAIR4HE) 88
2.5.5. Інституційне впровадження: політики, інфраструктура, бар'єри, показники результативності (KPI) 89

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ДО РОЗДІЛУ 2 92

ВИСНОВКИ 96

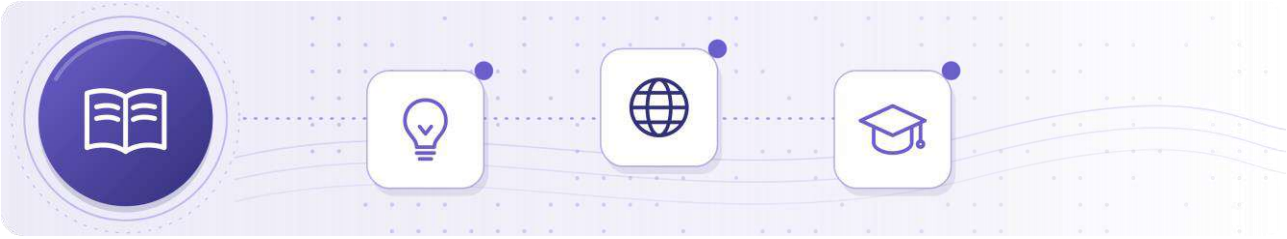
ГЛОСАРІЙ 100

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ



ЗВО – заклад вищої освіти
ЄС – Європейський Союз
ІЦО НАПН України – Інститут цифровізації освіти НАПН України
МОН України – Міністерство освіти і науки України
НАПН України – Національна академія педагогічних наук України
НДР – науково-дослідна робота
НРАТ – Національний репозитарій академічних текстів
УкрІНТЕІ – Український інститут науково-технічної експертизи та інформації
CC – Creative Commons (система відкритих ліцензій)
DMP – Data Management Plan (план управління даними)
DOI – Digital Object Identifier (цифровий ідентифікатор об’єкта)
EOSC – European Open Science Cloud (Європейська хмара відкритої науки)
FAIR – Findable, Accessible, Interoperable, Reusable (відшукуваність, доступність, сумісність, багаторазовість використання)
GDPR – General Data Protection Regulation (Загальний регламент ЄС про захист даних)
ISCED – International Standard Classification of Education (Міжнародна стандартна класифікація освіти)
ORCID – Open Researcher and Contributor ID (постійний ідентифікатор дослідника)
PID – Persistent Identifier (постійний ідентифікатор)
RDA – Research Data Alliance (Альянс дослідницьких даних)
ROR – Research Organization Registry (реєстр ідентифікаторів наукових організацій)

ПЕРЕДМОВА



Сучасний розвиток науки та освіти супроводжується стрімким зростанням обсягів дослідницьких даних, урізноманітненням їх типів і джерел, поширенням цифрових технологій та посиленням вимог до відкритості, прозорості й відтворюваності наукових досліджень. У таких умовах ефективне управління дослідницькими даними стає невід'ємною складовою сучасної наукової та освітньої діяльності. Особливого значення це набуває в галузі освітніх наук, де результати досліджень часто ґрунтуються на даних про освітні процеси, здобувачів освіти, педагогічних і науково-педагогічних працівників та інших учасників освітньої діяльності.

Одним із сучасних підходів до організації дослідницьких даних є концепція ***FAIR***, відповідно до якої дані мають бути *віднайденими (Findable*), *доступними (Accessible), сумісними (Interoperable)* та *придатними до багаторазового використання (Reusable)*. Реалізація цих принципів сприяє підвищенню якості та ефективності наукових досліджень, забезпечує можливість перевірки й відтворення отриманих результатів, створює умови для вторинного аналізу та інтеграції даних з різних джерел. Водночас FAIR не означає безумовної відкритості всіх даних: у разі наявності правових, етичних або інших обмежень мають бути забезпечені належне документування даних, їх відшукуваність та регламентований доступ.

Актуальність упровадження принципів FAIR у галузі освітніх наук зумовлена також необхідністю адаптації загальних підходів до специфіки педагогічних досліджень. Освітні наукові дані можуть мати кількісний, якісний або змішаний характер, містити персональну чи чутливу інформацію, потребувати спеціальних умов зберігання, захисту та доступу. Тому їх формування, опис, стандартизація, зберігання, публікація та повторне використання потребують урахування не лише технологічних, а й методологічних, правових та етичних вимог.

В Україні питання належного управління науковими даними набуває дедалі більшого значення у контексті розвитку відкритої науки та інтеграції

української наукової системи до європейського дослідницького простору. Важливим кроком у цьому напрямі стало затвердження Міністерством освіти і науки України у 2024 році Методичних рекомендацій щодо управління науковими даними для закладів вищої освіти (ЗВО) та наукових установ. Водночас для освітніх наук актуальним залишається розроблення та систематизація підходів до практичного застосування принципів FAIR з урахуванням особливостей педагогічних досліджень.

Довідник «Формування та використання FAIR-даних у галузі освітніх наук» підготовлено з метою систематизації теоретичних і практичних положень щодо організації роботи з дослідницькими даними відповідно до принципів FAIR та їх адаптації до потреб освітніх наук. Матеріали довідника охоплюють повний цикл роботи з даними – від їх планування, формування та документування до зберігання, публікації, пошуку, доступу, інтеграції, аналізу та повторного використання.

У *першому розділі* довідника розглянуто концептуальні засади FAIR, типи та класифікацію освітніх наукових даних, їх життєвий цикл і питання управління даними. Висвітлено підходи до стандартизації даних і метаданих, розроблення планів управління даними (Data Management Plan, DMP), вибору форматів і цифрових сервісів для зберігання та публікації даних. Окрему увагу приділено доступу, ліцензуванню, ідентифікації дослідницьких об’єктів, захисту персональних даних, анонімізації та псевдонімізації, а також можливостям використання сучасної наукової інфраструктури, зокрема репозитаріїв та Європейської хмари відкритої науки (European Open Science Cloud, EOSC).

Другий розділ присвячено використанню вже оприлюднених FAIR-даних у дослідницькій діяльності. Розглянуто стратегії пошуку та доступу до даних, оцінювання їх якості, правових умов і приватності, технічну та семантичну інтеграцію, цитування, забезпечення відтворюваності та фіксацію походження даних. Значну увагу приділено вторинному використанню даних, освітній аналітиці, кількісному й якісному аналізу, метааналізу, систематичним оглядам та іншим напрямам дослідницької діяльності.

Зміст довідника ґрунтується на аналізі сучасних наукових публікацій, міжнародних і вітчизняних документах, стандартах та практиках управління дослідницькими даними, а також на результатах науково-дослідної роботи (НДР) «Технології формування і використання FAIR даних в галузі освітніх наук» (ДР № 0126U000521). Представлені положення поєднують теоретичні засади з практичними рекомендаціями та прикладами, що дає змогу

використовувати довідник як орієнтир для організації роботи з даними на різних етапах освітнього дослідження.

Довідник адресовано науковим і науково-педагогічним працівникам, аспірантам, докторантам, здобувачам освіти, фахівцям наукових установ і ЗВО, бібліотечним та інформаційним працівникам, а також усім, хто займається формуванням, управлінням, зберіганням, публікацією, аналізом і повторним використанням дослідницьких даних у галузі освітніх наук. Довідник сприятиме поширенню принципів FAIR у дослідницькій практиці, підвищенню якості управління освітніми науковими даними та розвитку відкритої науки в Україні.

РОЗДІЛ І

ФОРМУВАННЯ FAIR-ДАНИХ У ГАЛУЗІ ОСВІТНІХ НАУК



1.1. Принципи FAIR: концептуальні засади



1.1.1. Виникнення та еволюція концепції FAIR

Формування концепції FAIR-даних є результатом тривалого еволюційного процесу, що охоплює понад два десятиліття розвитку ідей відкритого доступу до наукових даних. Передумовою виникнення принципів FAIR стали глобальні тенденції руху Open Access, що сформувалися на початку 2000-х років. Спочатку відкритий доступ стосувався переважно наукових публікацій: його засадничі положення були закладені Будапештською ініціативою відкритого доступу 2002 р. [1] та Берлінською декларацією про відкритий доступ до знань у галузі природничих і гуманітарних наук 2003 р. [2]. Обидва документи задекларували принцип вільного й безоплатного доступу до результатів наукових досліджень через мережу Інтернет, однак їхній фокус був зосереджений насамперед на публікаціях, а не на дослідницьких даних як таких. Поступово ідеї відкритості поширилися й на самі дані досліджень, що зумовило потребу у формулюванні окремих принципів управління ними.

Стрімке зростання обсягів дослідницьких даних, їх розпорошеність між різними платформами й системами, а також гостра потреба у верифікованості

наукових результатів актуалізували завдання створення уніфікованих, незалежних від конкретної дисципліни підходів до управління науковими даними. Відповіддю на ці виклики стала публікація у 2016 р. статті міжнародного колективу дослідників на чолі з М. Д. Вілкінсоном (M. D. Wilkinson), у якій уперше систематично викладено принципи FAIR – Findable (відшукуваність), Accessible (доступність), Interoperable (сумісність), Reusable (багаторазовість використання) – як узгоджений набір вимог до наукових даних і супровідних метаданих [3]. Новизна запропонованого підходу полягала у переорієнтації управління даними «від людини до машини»: принципи FAIR розроблялися не для зручності людини, а для автоматизованого машинного опрацювання, що передбачає розбудову інфраструктури репозитаріїв із постійними ідентифікаторами, машинозчитуваних метаданих, прикладних програмних інтерфейсів (API) та онтологій.

Економічне обґрунтування необхідності практичного впровадження принципів FAIR надало дослідження Європейської комісії: за оцінками аналітичного звіту 2018 р., щорічні втрати економіки Європейського Союзу (ЄС) через неналежне управління дослідницькими даними, тобто через недотримання принципів FAIR, перевищували 11 млрд євро [4]. Це підтвердило, що недотримання принципів FAIR спричиняє не лише наукові й репутаційні, а й відчутні фінансові втрати, що суттєво прискорило інституційне визнання FAIR як стандарту роботи з науковими даними на європейському рівні.

Після 2016 р. розвиток концепції FAIR відбувався за кількома взаємопов'язаними напрямами. *По-перше,* відбулася концептуальна стабілізація терміну: наукова спільнота дійшла згоди, що FAIR – це не конкретний технічний формат і не синонім «відкритості» даних, а набір властивостей цифрових об’єктів, що підвищують їхню придатність до повторного використання незалежно від того, чи є дані відкритими, чи обмеженого доступу. *По-друге,* сформувався запит на вимірюваність FAIR-відповідності: у 2019-2020 рр. робоча група Альянсу дослідницьких даних (Research Data Alliance, RDA) розробила модель FAIR-зрілості (FAIR Maturity Model) та чотирифазну методологію її оцінювання, результатом якої стали рекомендації, придатні до масового застосування й порівняльного аналізу [5]. *По-третє,* важливим інституційним кроком стала публікація Європейською комісією у 2018 р. звіту «Turning FAIR into Reality», який визначив дорожню карту практичної реалізації принципів у наукових екосистемах Європи, зокрема через розвиток Європейської хмари відкритої науки (EOSC) та ініціативи GO FAIR [6].

Освітня наукова спільнота розпочала активне обговорення принципів FAIR із 2018-2019 рр. Попри те, що FAIR спочатку розроблялися для природничих і точних наук – біомедицини, геноміки, кліматології, де вже існували усталені практики роботи з великими масивами даних, їх поступова адаптація до соціогуманітарних дисциплін, зокрема галузі освітніх наук, стала невідворотною тенденцією. Цей процес виявився складнішим через низку специфічних перешкод: гетерогенність освітніх даних, наявність чутливої персональної інформації про учасників освітнього процесу, відсутність усталених галузевих стандартів метаданих та розмаїття форматів подання освітніх ресурсів [7]. Актуальність цих питань засвідчують і дискусії у вітчизняній науковій спільноті щодо переходу до відкритої науки та впровадження принципів FAIR на європейському й національному рівнях [8].

У глобальному вимірі принципи FAIR отримали додаткове підкріплення в Рекомендаціях ЮНЕСКО щодо відкритої науки 2023 р., які визначили прозорість, доступність і повторне використання результатів досліджень одними з базових принципів відкритої науки на світовому рівні [9].

В Україні питання формування й використання FAIR-даних активно розробляється в контексті розвитку відкритої науки, інституційних репозитаріїв та підготовки методичних рекомендацій для ЗВО й наукових установ – зокрема, у працях Т. Ярошенко, С. Чуканової [10], С. Назаровця [11] та інших вітчизняних дослідників. Важливим кроком інституціоналізації принципів FAIR на національному рівні стало затвердження Міністерством освіти і науки України у 2024 р. Методичних рекомендацій щодо управління науковими даними для ЗВО та наукових установ, які вперше на нормативному рівні закріпили вимоги до планування, збереження та повторного використання наукових даних [12].

Отже, еволюція концепції FAIR пройшла шлях від загальних ідей відкритого доступу початку 2000-х років через формулювання чітких принципів управління науковими даними у 2016 р. до створення інструментів їх вимірювання та інституційного закріплення на європейському й глобальному нормативному рівнях. Для галузі освітніх наук ця еволюція триває й нині: інтеграція FAIR-концепції в освітній науковий дискурс відбувається поступово та нерівномірно, що зумовлює потребу подальшого теоретичного осмислення й адаптації принципів FAIR до специфіки педагогічних досліджень [7].

1.1.2. Принцип Findable (відшукуваність)

Принцип **Findable (відшукуваність)** є першим і визначальним у структурі FAIR-принципів, оскільки саме він забезпечує саму можливість виявлення дослідницьких даних людиною або автоматизованою системою – без чого доступність, сумісність і повторне використання даних втрачають сенс. Відповідно до базового формулювання принципів FAIR, відшукуваність декомпозується на чотири складники: (F1) дані та метадані отримують глобально унікальний і постійний ідентифікатор; (F2) дані описуються повними метаданими; (F3) метадані містять ідентифікатор даних, які вони описують; (F4) дані та метадані реєструються або індексуються у пошуковому ресурсі [3].

У розробленій моделі стандартизації FAIR-даних для галузі освітніх наук [13] відшукуваність перевіряють п'ятьма показниками, які дослідник може оцінити й самостійно: чи має набір даних постійний ідентифікатор (Digital Object Identifier (DOI), Handle), а не лише звичайне вебпосилання; чи достатньо повно його описано метаданими – від короткого опису за Дублінським ядром [14] до повного опису з контекстом дослідження, рівнем освіти та методологією; чи пов'язано опис безпосередньо з файлами даних; чи може набір даних знайти пошукова система; чи потрапляє опис до міжнародних каталогів наукових даних (DataCite, OpenAIRE). Докладніше показники розглянуто в підрозділі 1.4.3.

Специфіка освітніх наук накладає на реалізацію принципу Findable додаткові вимоги: значна частина освітніх даних містить персональну інформацію про учасників освітнього процесу, тому постійний ідентифікатор і відкриті метадані мають описувати набір даних, не розкриваючи водночас чутливого змісту самих даних, доступ до яких може бути обмеженим [15; 16].

1.1.3. Принцип Accessible (доступність)

Наступний принцип **Accessible (доступність)** визначає порядок отримання даних і метаданих після того, як їх було знайдено відповідно до принципу Findable. Згідно з базовим формулюванням FAIR-принципів, доступність декомпозується на два складники: (A1) дані отримуються за їхнім ідентифікатором через стандартизований комунікаційний протокол, який є відкритим, безкоштовним і загальнодоступним для реалізації, а за потреби передбачає процедуру автентифікації й авторизації; (A2) метадані залишаються доступними навіть тоді, коли самі дані більше не є доступними [3]. У розробленій моделі цей принцип представлено трьома індикаторами: A1 розділено на A-01-SP (доступність через стандартизований протокол –

HTTP/HTTPS, REST API) та A-02-AA (чіткість процедури автентифікації/авторизації – від автоматизованої, через ORCID (Open Researcher and Contributor ID (постійний ідентифікатор дослідника)) чи SSO, до ручного запиту); A2 реалізується через A-03-MA (збереженість метаданих після видалення даних) [13].

Принципове уточнення, яке супроводжує цей принцип від самого початку його формулювання, полягає в тому, що доступність не тотожна відкритості: дані можуть бути захищені чи мати обмежений доступ за умови, що процедура й умови цього доступу чітко й прозоро описані в метаданих, які, своєю чергою, залишаються відкритими завжди [17].

Протокольний рівень реалізації принципу Accessible забезпечується стандартизованими відкритими протоколами передавання даних – насамперед HTTP/HTTPS, а також FTP і протоколами прикладних програмних інтерфейсів (REST API). Використання саме відкритого й загальнодоступного протоколу є визначальною умовою автоматизованого отримання даних як людиною, так і програмою за прямим PID-посиланням. У цифрових репозитаріях, характерних для галузі освітніх наук (Zenodo, Figshare, Open Science Framework), доступ до пакета даних, як правило, гнучко налаштовується за кількома рівнями: відкритий доступ, обмежений доступ, закритий доступ з відкритими метаданими та доступ за ембарго – тобто з тимчасовим обмеженням, що знімається автоматично після завершення визначеного терміну [18]. Такий диференційований підхід дає змогу поєднати вимогу відкритості наукових даних із необхідністю тимчасового чи постійного обмеження доступу до чутливої інформації.

Ліцензування є юридичним інструментом, що унормовує умови фактичного використання даних після отримання доступу до них. Відсутність ліцензії за замовчуванням означає дію повного авторського права й забороняє будь-яке використання даних третіми особами, тоді як присвоєння стандартної відкритої ліцензії, найчастіше Creative Commons Attribution 4.0 International (CC BY 4.0) чи ліцензії публічного надбання CC0, знімає правову невизначеність і прямо в машинозчитуваних метаданих (а не лише текстом у файлі README) визначає дозволені способи використання набору даних [19]. Проміжний варіант – обмежені ліцензії (CC BY-NC, CC BY-ND) чи угоди про використання даних (Data Use Agreement) – застосовується тоді, коли повна відкритість неможлива з етичних чи правових причин, однак принаймні метадані та умови доступу залишаються задокументованими й відшукуваними.

Автентифікація й авторизація забезпечують контрольований доступ до даних обмеженого поширення, не порушуючи водночас вимоги відшукуваності метаданих. У цифрових репозитаріях розрізняють автоматизовані процедури (реєстрація через ORCID, інституційний єдиний вхід (SSO), токен доступу до API) та ручні процедури (запит доступу електронною поштою до розпорядника даних, розгляд заявки комітетом з етики). Перший варіант відповідає вищому рівню відповідності принципу Accessible, оскільки дає змогу автоматично й відтворювано підтвердити право на доступ, тоді як другий передбачає участь людини на кожному етапі й уповільнює процес повторного використання даних [13].

Специфіка освітніх наук зумовлює підвищені вимоги до реалізації принципу Accessible, оскільки значна частина освітніх даних стосується учасників освітнього процесу, у тому числі неповнолітніх, і належить до персональних даних у контексті Закону України «Про захист персональних даних» від 01.06.2010 № 2297-VI та Загального регламенту про захист даних Європейського Союзу (General Data Protection Regulation, GDPR) [20; 21]. Практичним рішенням, яке узгоджує принцип відкритості з вимогами законодавства, є розмежування рівнів доступу: метадані набору даних (опис вибірки, методології, інструментарію) залишаються відкритими й відшукуваними завжди, тоді як доступ до самих первинних даних обмежується анонімізацією чи псевдонімізацією, отриманням інформованої згоди учасників дослідження на повторне використання, а за потреби і процедурою автентифікації дослідника, який звертається за доступом [15; 18].

Таким чином, реалізація принципу Accessible в галузі освітніх наук передбачає узгоджену дію трьох складників – відкритого комунікаційного протоколу для технічного отримання даних, прозорої ліцензії для правового визначення умов їх використання та належно налаштованої процедури автентифікації для контрольованого доступу до чутливих даних, що в сукупності забезпечує баланс між відкритістю наукових даних і захистом прав учасників освітнього процесу. Цей баланс є передумовою для реалізації наступного принципу – Interoperable, який визначає технічну сумісність уже отриманих даних із іншими наборами даних та інструментами.

1.1.4. Принцип Interoperable (сумісність)

Принцип Interoperable (сумісність, або інтероперабельність) визначає вимоги до технічної й семантичної сумісності даних, отриманих і збережених відповідно до принципів Findable та Accessible, з іншими наборами даних, програмними інструментами й аналітичними системами. Відповідно до

базового формулювання FAIR-принципів, інтероперабельність декомпозується на три складники: (I1) дані та метадані використовують формальну, доступну, спільну й широко застосовувану мову подання знань; (I2) дані та метадані використовують словники, що самі відповідають принципам FAIR; (I3) дані та метадані містять кваліфіковані посилання на інші дані та метадані [3]. Модель зіставляє їх із трьома індикаторами: I-01-KL (від семантичних форматів RDF/JSON-LD/OWL [22; 23] до структурованих CSV/JSON/XML чи неструктурованих PDF/DOCX); I-02-LS (використання контрольованих словників із глобальними ідентифікаторами – International Standard Classification of Education (ISCED), EuroVoc – проти довільної термінології); I-03-QR (наявність активних ідентифікаторів пов'язаних ресурсів) [13].

У практичній реалізації розрізняють два рівні інтероперабельності – синтаксичний і семантичний. Синтаксична сумісність забезпечується використанням загальновживаних структурованих форматів подання даних (CSV, TSV, JSON, XML), які може зчитати широкий спектр програмного забезпечення, однак без додаткового опису семантики полів. Семантична сумісність – це вищий рівень, за якого дані подаються у форматах, що прямо кодують значення й зв'язки між елементами: RDF (Resource Description Framework) [22] та мова опису онтологій OWL (Web Ontology Language) [23], а також похідний від них формат серіалізації JSON-LD. Саме семантична сумісність дає змогу автоматично зіставляти значення змінних з різних досліджень, розпізнавати еквівалентні поняття, подані різними термінами, та інтегрувати дані кількох наборів у єдиний аналітичний контекст без участі людини, оскільки, як показано в дослідженнях семантичної інтероперабельності, вона не може бути досягнута без опори на формальні онтологічні теорії [24].

Ключову роль у забезпеченні семантичної сумісності відіграють онтології та контрольовані словники – формалізовані системи понять певної предметної галузі зі стійкими ідентифікаторами кожного терміна. Для галузі освітніх наук такими словниками є Міжнародна стандартна класифікація освіти ISCED (ЮНЕСКО), яка дає змогу однозначно позначати рівень і галузь освіти, багатомовний тезаурус термінології європейської публічної політики EuroVoc, а також вузькогалузеві освітні тезауруси й словники метаданих, побудовані відповідно до ДСТУ ISO 25964-1:2018. На відміну від довільної термінології чи локальних абревіатур, використання таких словників, кожен елемент яких має власний глобальний ідентифікатор, дозволяє однозначно зіставляти терміни між дослідженнями різних авторів, установ і країн [25].

Опис самого набору даних для потреб міжсистемного обміну забезпечують стандартизовані схеми метаданих. Дублінське ядро (Dublin Core) залишається базовим рівнем опису, придатним для більшості цифрових об'єктів, однак не охоплює галузевої специфіки. Для соціогуманітарних і освітніх досліджень, що ґрунтуються на опитуваннях, анкетуванні й педагогічних експериментах, спеціалізованим стандартом метаданих є DDI (Data Documentation Initiative), який детально описує змінні, кодові книги, вибірку та методику збору даних [26]. Для даних, що безпосередньо розміщені на освітніх платформах і в системах управління навчанням, застосовуються вузькогалузеві стандарти: IEEE 1484.12.1 – стандарт метаданих навчальних об'єктів (Learning Object Metadata, LOM) [27], та xAPI (Experience API) Ініціативи передових розподілених технологій навчання (ADL) – специфікація для фіксації та обміну даними про навчальний досвід і активність здобувачів освіти в уніфікованому форматі JSON-тверджень «суб'єкт – дія – об'єкт» [28].

Для потреб вимірюваного оцінювання відповідності освітніх даних принципу Interoperable авторами розробленої моделі стандартизації FAIR-даних для галузі освітніх наук запропоновано три індикатори з трибальною шкалою (0-2 бали): використання формальної мови подання знань (від семантичних форматів RDF, JSON-LD, OWL до загальновживаних структурованих форматів CSV, JSON, XML та неструктурованих чи закритих форматів PDF, DOCX, сканованих копій) і використання стандартизованих словників, що відповідають принципам FAIR (від контрольованих вокабулярів і онтологій із глобальними ідентифікаторами, як ISCED і EuroVoc, до загальноприйнятих термінів без посилання на класифікатори або довільної термінології) і наявність кваліфікованих посилань на пов'язані ресурси з активними ідентифікаторами [13]. Найвищий рівень відповідності моделі (Max-FAIR) передбачає подання набору даних у форматі RDF або JSON-LD з використанням онтології ISCED та вузькогалузевих стандартів xAPI чи LOM, що робить його придатним до автоматизованого машинного аналізу [13].

Таким чином, реалізація принципу Interoperable у галузі освітніх наук передбачає поступовий перехід від синтаксичної сумісності структурованих форматів до семантичної сумісності на основі формальних мов подання знань, контрольованих словників і галузевих стандартів метаданих. Забезпечена в такий спосіб технічна й термінологічна узгодженість даних є безпосередньою передумовою реалізації останнього принципу FAIR – Reusable, який визначає умови їх повторного, зокрема вторинного й метааналітичного, використання в педагогічних дослідженнях.

1.1.5. Принцип Reusable (придатність до повторного використання)

Принцип Reusable (багаторазовість використання або придатність до повторного використання) завершує структуру FAIR-принципів і визначає умови, за яких знайдені, доступні й сумісні дані можуть бути коректно застосовані іншими дослідниками для нових наукових цілей. Відповідно до базового формулювання, повторне використання декомпозується на чотири складники: (R1) дані та метадані описані численними точними й релевантними атрибутами; (R1.1) дані супроводжуються зрозумілою й доступною ліцензією на використання; (R1.2) дані супроводжуються детальним описом походження (provenance); (R1.3) дані та метадані відповідають релевантним для галузі стандартам спільноти [3]. Пізніше опубліковане тлумачення цих складників уточнило, що вимоги R1.1–R1.3 фактично є конкретизацією загальної вимоги R1: ліцензія, походження та галузеві стандарти – це саме ті категорії метаданих, повнота й точність яких визначають практичну придатність набору даних до повторного використання [29].

Розроблена авторами модель стандартизації зіставляє їх із чотирма індикаторами: R1→R-01-RA (повнота опису вибірки, методики, інструментарію); R1.1→R-02-LA (наявність стандартної ліцензії в метаданих, а не лише в README); R1.2→R-03-DP (документальне підтвердження усіх етапів обробки, в ідеалі – зі скриптами; технічна основа – онтологія PROV-O); R1.3 → R-04-DS (відповідність xAPI, ISCED, LOM) [13].

Специфіка галузі освітніх наук зумовлює додаткові обмеження повторного використання: інформована згода учасників дослідження, отримана на етапі збирання даних, як правило, окреслює конкретну мету використання, тому вторинний аналіз чужих освітніх даних, особливо тих, що стосуються неповнолітніх, потребує перевірки, чи охоплює первинна згода такий спосіб повторного застосування, а за потреби – додаткової анонімізації чи псевдонімізації [7]. Розвиток у дослідників компетентності роботи з FAIR-даними – планування, документування, ліцензування та подальше повторне використання – визнано окремим системним завданням: у межах європейського проєкту FAIRsFAIR розроблено Рамку компетентностей FAIR для вищої освіти (FAIR Competence Framework for Higher Education, FAIR4HE), яка визначає рівні володіння цими навичками для бакалаврського, магістерського та докторського рівнів підготовки [30].

Разом чотири принципи FAIR – Findable, Accessible, Interoperable, Reusable формують цілісний цикл управління освітніми дослідницькими даними: від їх виявлення через отримання доступу й забезпечення технічної сумісності до

коректного повторного застосування в нових наукових і практичних контекстах.

1.2. Типи освітніх наукових даних та їх класифікація



Освітні наукові дані – це будь-яка інформація, що збирається, реєструється або генерується в процесі педагогічного дослідження і слугує підставою для формулювання наукових висновків. Розуміння типології таких даних є відправною точкою для прийняття рішень щодо форматів зберігання, стандартів метаданих і режимів доступу, розглянутих у наступних підрозділах довідника (1.4, 1.6, 1.8): саме тип даних значною мірою визначає, які стандарти метаданих, формати файлів і рівні захисту застосовні до конкретного набору. У межах виконання НДР «Технології формування і використання FAIR даних в галузі освітніх наук» (ДР № 0126U000521) розроблено багатовимірну таксономію освітніх наукових даних [1], основними складниками якої є кількісні, якісні та змішані дані, отримані безпосередньо в дослідженні, адміністративні й інституційні освітні дані, що існують незалежно від конкретного дослідницького проєкту, а також дані, класифіковані за типом і методом дослідження, у межах якого їх було отримано.

1.2.1. Кількісні дані

Кількісні дані подаються у числовій формі й допускають статистичну обробку. У педагогічних дослідженнях до них належать: результати стандартизованих тестів і контрольних зрізів успішності; статистичні показники успішності – оцінки, бали, рейтинги, відсоток виконання завдань; дані моніторингових досліджень освіти, зокрема результати зовнішнього незалежного оцінювання та міжнародних порівняльних досліджень (PISA, TIMSS); показники відвідуваності, успішності та відрахування; результати анкетування за шкалами Лайкерта; журнали активності в електронних навчальних середовищах (LMS-логи). Такі дані, як правило, зберігаються у форматах CSV, XLSX, SPSS (.sav), R-форматах або структурованих базах даних SQL.

Для опису кількісних освітніх даних рекомендується стандарт метаданих DDI (Data Documentation Initiative), спеціально розроблений для соціальних і поведінкових наук і широко застосовуваний в архівах освітніх даних, зокрема ICPSR та CESSDA [2]. Профіль DDI Codebook передбачає докладний опис кожної змінної (назву, формулювання питання, шкалу вимірювання, допустимі значення, правила кодування пропущених значень), що є визначальною умовою відповідності таких даних принципу Reusable: без кодової книги статистичний масив цифр залишається практично непридатним до коректної повторної інтерпретації іншим дослідником, навіть якщо технічний доступ до файлу відкритий.

1.2.2. Якісні дані

Якісні дані охоплюють текстові, аудіо- та відеоматеріали, що вимагають інтерпретаційного аналізу. У педагогічних дослідженнях до них належать: транскрипти інтерв'ю та фокус-групових дискусій; записи педагогічних спостережень і польові нотатки; відеозаписи уроків та навчальних занять; документи (навчальні програми, плани уроків, учнівські роботи); матеріали відкритих запитань анкет. Зберігання якісних даних потребує особливої уваги до вибору формату, оскільки різні формати мають різний потенціал довготривалого зберігання: TXT, RTF або PDF/A рекомендовані для текстових матеріалів; WAV чи FLAC – для аудіозаписів інтерв'ю та спостережень; MP4 (кодек H.264) – для відеозаписів занять.

Для метаданих якісних даних принципове значення має опис застосованого методу аналізу – тематичного кодування [3], заземленої теорії, дискурс-аналізу та програмних засобів, що використовувалися для обробки (MAXQDA, NVivo, ATLAS.ti). Аналогічно до кодової книги для кількісних даних, для якісних наборів роль документа, що забезпечує повторне використання, відіграє протокол аналізу: він фіксує кодову схему, категорії й підкатегорії тем, а також процедури забезпечення надійності інтерпретації (зокрема, участь кількох незалежних кодувальників), без яких транскрипт інтерв'ю чи відеозапис заняття залишається складним для верифікації й повторного застосування іншими дослідниками.

Практичною проблемою якісних досліджень є те, що результати кодування, виконаного в одній програмі (MAXQDA, NVivo, ATLAS.ti), раніше неможливо було відкрити в іншій. Для її розв'язання розробники цих програм спільно створили відкритий стандарт обміну REFI-QDA, який дає змогу переносити між програмами проєкт разом із джерелами, виділеними

фрагментами та кодами [4]. Для освітян, які не мають доступу до комерційних програм, існують безкоштовні інструменти з відкритим кодом, зокрема Taguette, що дає змогу імпортувати тексти, позначати й кодувати фрагменти та експортувати результати у відкритих форматах [5]. Під час архівування результатів кодування бажано зберігати їх у форматі REFI-QDA або хоча б експортувати кодову схему в текстовому чи табличному форматі.

1.2.3. Змішані типи даних

Більшість сучасних педагогічних досліджень використовують змішані методи, що передбачають поєднання кількісних і якісних даних у межах одного дослідження [6]. Типовими прикладами є анкетування, що поєднує закриті питання (кількісна складова, придатна до статистичної обробки) з відкритими питаннями (якісна складова, що потребує змістового аналізу відповідей), а також кейс-стаді (дослідження випадку), у межах якого кількісні показники успішності учасника чи групи доповнюються якісними даними спостережень, інтерв'ю з учасниками та документальними матеріалами. Набір даних педагогічного експерименту змішаного типу може, наприклад, включати результати тестів (кількісна складова) та транскрипти спостережень за перебігом експерименту (якісна складова), пов'язані спільним ідентифікатором респондента чи етапу педагогічного впливу.

Підготовка змішаних наборів даних до публікації в репозитарії потребує особливої уваги до внутрішньої узгодженості пакета: необхідно забезпечити чітку документацію зв'язків між кількісними й якісними компонентами (наприклад, через спільний ідентифікатор учасника чи хронологічну мітку етапу дослідження), окремий опис методики й формату кожного типу файлів та єдину, послідовну систему іменування файлів у межах усього набору. Відсутність такої узгодженості є типовою причиною, через яку змішані набори освітніх даних, попри формальну відповідність окремим FAIR-показникам своїх компонентів, залишаються практично непридатними до цілісного повторного використання.

Організація змішаного набору даних значною мірою залежить від обраного дизайну дослідження. У методології змішаних досліджень розрізняють, зокрема, конвергентний дизайн (кількісні й якісні дані збирають паралельно і зіставляють під час інтерпретації), пояснювальний послідовний дизайн (спершу збирають кількісні дані, а потім якісні – для пояснення отриманих результатів) і дослідницький послідовний дизайн (спершу якісний етап, на основі якого розробляють інструмент для кількісного етапу) [6]. Для FAIR-даних це має

практичне значення: у файлі README варто зазначити дизайн дослідження й послідовність етапів, щоб інший дослідник розумів, які дані є первинними, а які отримано на їх основі.

1.2.4. Адміністративні та інституційні освітні дані

За іншою класифікаційною ознакою – походженням і первинним призначенням розрізняють первинні дослідницькі дані, зібрані безпосередньо для розв'язання завдань дослідження, та адміністративні й інституційні дані освітніх установ: реєстри учнів і студентів, дані про склад педагогічного персоналу, фінансова звітність закладів освіти, результати інституційних моніторингів якості освіти. На відміну від первинних дослідницьких даних, адміністративні дані є вторинними для дослідника: вони збираються органами управління освітою чи самими закладами освіти для управлінських, а не дослідницьких цілей, і лише згодом залучаються до наукового аналізу. За своєю природою адміністративні дані можуть бути кількісними (реєстри, статистичні показники, результати оцінювання) або якісними (протоколи, звіти, накази, текстові характеристики), тому на них поширюються вимоги до документування, описані в підрозділах 1.2.1–1.2.3.

Адміністративні освітні дані, як правило, перебувають у власності закладів освіти й органів управління освітою та мають суттєві обмеження доступу через вимоги конфіденційності: значна їх частина підпадає під дію Закону України «Про захист персональних даних» від 01.06.2010 № 2297-VI [7], оскільки прямо чи опосередковано стосується конкретних здобувачів освіти або працівників закладу. Для дослідницького використання такі дані, як правило, надаються у знеособленому чи агрегованому вигляді – на рівні класу, закладу, регіону чи країни, а не окремого учня чи вчителя. Прикладами публічно доступних адміністративних освітніх баз даних, що вже пройшли необхідну агрегацію й публікуються офіційними інституціями, є дані Інституту статистики ЮНЕСКО щодо освіти у світі (UNESCO Institute for Statistics) [8], статистика Організації економічного співробітництва та розвитку «Погляд на освіту» (OECD, Education at a Glance) [9] та відкриті дані Міністерства освіти і науки України.

На відміну від дослідницьких кількісних, якісних і змішаних даних, для яких дослідник самостійно визначає стандарт метаданих і рівень документування ще на етапі планування збору, адміністративні дані здебільшого успадковують структуру й формат, визначені інформаційною системою органу управління освітою чи закладу освіти. Це означає, що дослідник, який залучає адміністративні дані до наукового аналізу, має

додатково документувати їх походження (джерело, дату вивантаження, застосовані процедури агрегації та анонімізації) відповідно до вимог принципу Reusable, оскільки сама інформаційна система-джерело, як правило, не надає таких метаданих у форматі, придатному для наукового цитування й повторного використання.

1.2.5. Класифікація за типами і методами дослідження

Розглянутий вище поділ на кількісні, якісні, змішані та адміністративні дані описує дані за їхньою природою й походженням, проте не характеризує сам тип наукового дослідження, у межах якого дані було отримано. Це самостійна, незалежна вісь класифікації, яку виокремлено в розробленій у межах дослідження таксономії FAIR-даних для наукових досліджень у галузі освіти на основі узагальнення підходів вітчизняного та зарубіжного методологічного доробку [1; 10] і яка охоплює сім груп даних.

Експериментальні та квазіекспериментальні дані отримують в умовах спеціально організованих педагогічних експериментів із повним або частковим дотриманням вимог щодо однорідності експериментальних і контрольних груп. Кореляційні та регресійні дані виявляють статистичні зв'язки між двома або більше змінними або моделюють значення залежної змінної як функцію одного чи кількох факторів впливу на освітній процес. Описові дані містять звіти національних, регіональних та інституційних зрізів якості освіти, дані зовнішнього оцінювання, моніторинг умов навчання. Дані якісних досліджень отримують методами інтерв'ю, фокус-груп, спостережень і кейсових досліджень – і за своєю природою значною мірою збігаються з якісними даними, розглянутими в підрозділі 1.2.2, проте виокремлюються тут саме як результат конкретного методу дослідження, а не за форматом самих даних.

Дані теоретичних досліджень – тип, не представлений серед розглянутих раніше категорій, – охоплюють моделі, експортовані набори метаданих статей, таксономії й онтології, що є результатом концептуального, а не емпіричного дослідження. Методологічні дані містять опис методик, цифрових інструментів і засобів оцінювання, застосованих у дослідженні. Дані, отримані через застосування змішаних методів дослідження, містять узгоджені набори кількісних і якісних даних зі спільними ідентифікаторами респондентів чи етапів педагогічного впливу – це той самий тип, що розглянутий у підрозділі 1.2.3, але позначений тут з погляду методу його отримання.

Ці дві осі класифікації – за природою даних (підрозділи 1.2.1-1.2.4) і за типом та методом дослідження (цей підрозділ) – не суперечать одна одній, а

застосовуються паралельно: наприклад, набір експериментальних даних водночас є кількісним, а набір даних якісного дослідження – якісним за форматом. Водночас категорія теоретичних даних не має прямого відповідника серед кількісних, якісних, змішаних чи адміністративних даних, що засвідчує необхідність її окремого врахування під час опису метаданими: модель, таксономія чи онтологія, яку дослідник публікує як результат теоретичного дослідження, потребує інших полів метаданих (насамперед посилань на вихідні джерела й пов'язані концепти), ніж емпіричний набір спостережень.

Отже, повний опис типології освітніх наукових даних вимагає врахування обох розглянутих осей класифікації одночасно: природи даних, що визначає формат файлів і стандарт метаданих, та типу й методу дослідження, що визначає структуру супровідної документації й контекст, необхідний для коректного повторного використання. Це зумовлює потребу в цілісній, вимірюваній моделі стандартизації FAIR-даних, що враховує специфіку кожного типу даних і водночас забезпечує їх порівняння у межах єдиної системи критеріїв.

1.3. Життєвий цикл даних і суб'єкти управління FAIR-даними



Формування FAIR-даних, розглянуте в підрозділах 1.1–1.2 на рівні принципів і типології, на практиці розгортається як послідовність конкретних дій багатьох учасників дослідницького процесу, розподілених у часі. Цей підрозділ розглядає дві взаємопов'язані складові такого практичного виміру: етапи життєвого циклу дослідницьких даних та ролі учасників, відповідальних за виконання дій на кожному етапі.

1.3.1. Етапи життєвого циклу дослідницьких даних

Життєвий цикл дослідницьких даних – це послідовність стадій, які проходить набір даних від задуму дослідження до його довготривалого архівування чи знищення. Найбільш поширеною референтною моделлю життєвого циклу є розроблена Digital Curation Centre (Велика Британія) DCC Curation Lifecycle Model, яка виокремлює дії повного життєвого циклу (опис і репрезентаційна інформація, планування збереження, моніторинг спільноти й

участь у ній, курування та збереження), послідовні дії (концептуалізація, створення чи отримання даних, оцінювання й відбір, внесення в систему зберігання (ingest), дії зі збереження, зберігання, доступ і повторне використання, трансформація) та періодичні дії (утилізація, повторне оцінювання, міграція), що застосовуються за потреби на різних стадіях [1].

Адаптований до специфіки педагогічного дослідження, життєвий цикл FAIR-даних охоплює шість практичних етапів. Планування – визначення типів даних (за типологією, розглянутою в підрозділі 1.2), методів збору та попереднього плану управління даними (підрозділ 1.5) ще до початку дослідження. Збирання й обробка – безпосереднє отримання даних (тестування, опитування, спостереження) та їх очищення, кодування, структурування. Опис і документування – присвоєння метаданих і постійних ідентифікаторів відповідно до стандартів, розглянутих у підрозділі 1.4, що реалізують принципи Findable й Interoperable. Зберігання та публікація – розміщення в репозитарії (підрозділ 1.7) з визначеним режимом доступу та ліцензією, що реалізують принципи Accessible й Reusable. Доступ і повторне використання – етап, детально розглянутий у розділі 2 довідника, на якому дані застосовуються іншими дослідниками. Архівування чи знищення – довготривале збереження набору даних за умови його наукової цінності або контрольоване видалення, якщо термін зберігання, визначений умовами згоди учасників дослідження (підрозділ 1.9), вичерпано.

Ключова відмінність життєвого циклу від лінійного процесу полягає в його ітеративності: дані можуть повертатися на етап повторного оцінювання чи трансформації формату (наприклад, під час підготовки до метааналізу, розглянутого в підрозділі 2.3), а рішення, ухвалені на етапі планування, безпосередньо визначають, наскільки успішно набір даних пройде наступні етапи – зокрема, недостатньо продумана структура метаданих на етапі опису ускладнює подальший пошук і повторне використання даних.

1.3.2. Розподіл ролей і відповідальності в управлінні дослідницькими даними

Ефективне управління FAIR-даними на кожному етапі життєвого циклу вимагає чіткого розподілу відповідальності між учасниками дослідницького колективу: практика управління науковими даними демонструє, що обов’язки з управління даними мають бути явно закріплені за конкретними особами, оскільки нечіткий розподіл ролей є типовою причиною втрати метаданих, незадокументованих трансформацій і затримок публікації [2].

Керівник дослідження (Principal Investigator, PI) несе загальну відповідальність за управління даними в межах проєкту: він затверджує план управління даними (підрозділ 1.5), забезпечує відповідність вимогам організації, що фінансує, та наукової установи, а також може делегувати окремі завдання іншим учасникам команди, зберігаючи кінцеву відповідальність за дотримання плану [2]. Data steward (фахівець з управління дослідницькими даними) – роль, що набула окремого визнання в міжнародній практиці управління дослідницькими даними: це фахівець, відповідальний за щоденне виконання зобов'язань плану управління даними – присвоєння метаданих, перевірку відповідності контрольованим словникам, забезпечення зв'язку між даними та їхнім описом, а також перевірку відповідності набору даних принципам FAIR перед його публікацією [3]. Підготовка фахівців цієї ролі визнана окремим освітнім завданням: зокрема, розроблено спеціалізовані навчальні програми з управління FAIR-даними для дослідників, які не мають професійної підготовки в галузі інформаційних наук [4].

Фахівець із захисту персональних даних, юрист чи дослідник з питань етики відповідає за узгодження практики роботи з даними відповідно до законодавства про захист персональних даних – Законом України № 2297-VI та Загальним регламентом про захист даних (GDPR), розглянутими в підрозділі 1.9, – зокрема за оцінювання обсягу інформованої згоди учасників дослідження, погодження меж анонімізації чи псевдонімізації та визначення дозволених способів повторного використання чутливих освітніх даних. IT-супровід забезпечує технічну інфраструктуру: інтеграцію з репозитарієм (підрозділ 1.7), налаштування процедур автентифікації й авторизації (підрозділ 1.8.2), резервне копіювання та технічну підтримку інструментів анонімізації (підрозділ 1.9.2).

У міжнародній практиці розподіл цих ролей формалізовано, зокрема, у таксономії ролей учасників дослідження CRediT (Contributor Roles Taxonomy), що з 2022 р. є офіційним стандартом ANSI/NISO Z39.104-2022: серед чотирнадцяти визначених у ній ролей окремо виділено «курування даних» (Data curation) – управлінську діяльність з анотування (створення метаданих), очищення й супроводу дослідницьких даних для первинного й подальшого використання [5]. Водночас, як засвідчує практика, у невеликих дослідницьких колективах, типових для педагогічних досліджень, одна й та сама особа нерідко поєднує кілька ролей одночасно – наприклад, керівник дослідження може виконувати й функції розпорядника даних, що не скасовує потреби чітко визначити й задокументувати, хто саме відповідає за кожну конкретну дію на кожному етапі життєвого циклу даних, розглянутому в підрозділі 1.3.1 [2].

1.4. Стандартизація FAIR-даних в освітніх науках



Стандартизація є технічною й організаційною передумовою практичної реалізації принципів FAIR, розглянутих раніше: без узгоджених форматів метаданих, спільної термінології та вимірюваних критеріїв оцінювання декларація відповідності набору освітніх даних принципам FAIR залишається суб'єктивним твердженням, яке неможливо ні перевірити, ні порівняти між дослідженнями. У цьому підрозділі послідовно розглянуто загальнонаукові стандарти управління даними, спеціалізовані освітні стандарти метаданих та розроблену в межах дослідження модель стандартизації FAIR-даних для галузі освітніх наук.

1.4.1. Міжнародні стандарти управління науковими даними

Dublin Core (DC) – розроблений Dublin Core Metadata Initiative (DCMI) мінімальний набір із п'ятнадцяти елементів метаданих (назва, автор, тема, опис, видавець, дата, тип, формат, ідентифікатор, джерело, мова, відношення, охоплення, права та учасник), що є найбільш поширеним і водночас найпростішим стандартом опису цифрового наукового об'єкта [1]. Його головна перевага – універсальність і сумісність практично з будь-яким типом ресурсу, а головне обмеження – відсутність галузевої деталізації, необхідної для повноцінного опису дослідницьких даних.

DataCite Metadata Schema – спеціалізований стандарт метаданих для наукових даних, що супроводжуються постійним ідентифікатором (найчастіше DOI). Схема визначає обов'язкові властивості (ідентифікатор, автор, назва, видавець, рік публікації, тип ресурсу) та рекомендовані властивості (предметні рубрики, суміжні ідентифікатори, права, розмір, версія, географічне охоплення), що безпосередньо забезпечують реалізацію принципів F та R [2]. Схема DataCite є стандартом за умовчанням для більшості наукових репозитаріїв, розглянутих у підрозділі 1.7 довідника, включно з Zenodo, Figshare та OSF.

Для позначення предметної галузі й термінології освітнього дослідження в цих схемах метаданих застосовуються також загальногалузеві контрольовані

словники, такі як Тезаурус ЮНЕСКО з питань освіти (UNESCO Thesaurus) та Тезаурус ERIC (Education Resources Information Center), що доповнюють розглянуту в підрозділі 1.1.4 класифікацію ISCED власне термінологічним, а не лише структурним рівнем контролю [3].

Schema.org – спільний словник структурованих даних, заснований 2011 р. провідними пошуковими системами (Google, Microsoft, Yahoo!) для розмітки вебсторінок у форматах JSON-LD, Microdata або RDFa [4]. На відміну від Dublin Core та DataCite, які призначені передусім для внутрішнього опису набору даних у репозитарії, Schema.org виконує функцію зовнішньої, орієнтованої на пошукові системи розмітки: розмічена сторінка стає розпізнаваною для Google Dataset Search та подібних сервісів як самостійний індексований об'єкт типу Dataset, що прямо реалізує складник F4 (індексація в пошуковому ресурсі) принципу Findable, розглянутого раніше. Актуальна версія словника Schema.org налічує понад вісімсот типів і півтори тисячі властивостей і використовується на десятках мільйонів вебдоменів, що робить його одним із найпоширеніших механізмів забезпечення машинної видимості наукових даних у відкритому вебпросторі.

1.4.2. Освітні метадані та стандарти

Поряд із загальнонауковими стандартами метаданих для галузі освітніх наук застосовуються спеціалізовані стандарти, що враховують педагогічну специфіку описуваних об'єктів. IEEE LOM (Learning Object Metadata, IEEE 1484.12.1) – міжнародний стандарт метаданих для опису навчальних об'єктів, що визначає дев'ять категорій елементів: загальну інформацію, дані життєвого циклу, метаінформацію, технічні характеристики, освітні характеристики (тип взаємодії, рівень складності, типовий вік користувача, педагогічний контекст), права, зв'язки з іншими ресурсами, анотації та класифікацію [5]. Саме категорія освітніх характеристик відрізняє LOM від загальнонаукових схем метаданих і робить його придатним для опису навчальних матеріалів, курсів і освітніх ресурсів, а не лише результатів дослідження в класичному розумінні.

1EdTech Consortium (до 2022 р. – IMS Global Learning Consortium) – некомерційний консорціум, заснований 1996 р., що розробляє специфікації інтероперабельності освітніх технологій, зокрема стандарти обміну даними між системами управління навчанням, платформами оцінювання та інструментами навчальної аналітики [6]. Найбільш затребуваними специфікаціями 1EdTech для освітніх наукових даних є QTI (Question and Test Interoperability) для обміну тестовими завданнями та результатами оцінювання, а також Caliper Analytics –

специфікація для збору й обміну даними про навчальну активність, що за призначенням є прямим аналогом розглянутого раніше стандарту xAPI. LRMI (Learning Resource Metadata Initiative) – спеціалізоване розширення словника Schema.org для опису освітніх ресурсів, розроблене спільно Creative Commons та Асоціацією освітніх видавців, стюардом якого з 2014 р. є Dublin Core Metadata Initiative [7]. LRMI додає до базових типів Schema.org освітньо специфічні властивості цільову аудиторію (audience), рівень освіти (educationalLevel), тип навчального ресурсу (learningResourceType) та відповідність освітнім стандартам (educationalAlignment), що дає змогу поєднати переваги масштабованої веброзмітки Schema.org із галузевою точністю опису, притаманною IEEE LOM.

Порівняння трьох розглянутих освітніх стандартів засвідчує їхню взаємодоповнюваність, а не конкуренцію: IEEE LOM забезпечує найбільш вичерпний, структурований опис навчального об'єкта для внутрішнього використання в системах управління навчанням; 1EdTech охоплює насамперед функціональну інтероперабельність між платформами (обмін тестами, результатами, даними активності); LRMI, своєю чергою, оптимізований для зовнішньої відшукуваності освітнього ресурсу в загальному вебпросторі завдяки інтеграції зі Schema.org. Для повноцінного опису набору освітніх наукових даних, що супроводжує педагогічне дослідження, доцільним є комбіноване застосування: DataCite чи Dublin Core для базової наукової атрибуції набору даних, LRMI-розмітка для його зовнішньої відшукуваності пошуковими системами, а специфікації 1EdTech чи формат xAPI для власне освітніх даних активності, якщо дослідження оперує журналами взаємодії здобувачів із цифровим навчальним середовищем.

1.4.3. Модель стандартизації FAIR-даних для галузі освітніх наук: рівні, критерії та показники оцінювання

На основі аналізу міжнародного досвіду та специфіки педагогічних досліджень у межах виконання НДР «Технології формування і використання FAIR даних в галузі освітніх наук» (ДР № 0126U000521) розроблено модель стандартизації FAIR-даних для галузі освітніх наук (рис. 1.1).

Модель складається з чотирьох блоків. *Цільовий блок* визначає мету – стандартизацію даних наукових досліджень у галузі освітніх наук відповідно до принципів FAIR та її підґрунтя: законодавство України й директиви ЄС, міжнародні стандарти, нормативні документи ЗВО й наукових установ, потреби наукових і науково-педагогічних працівників.

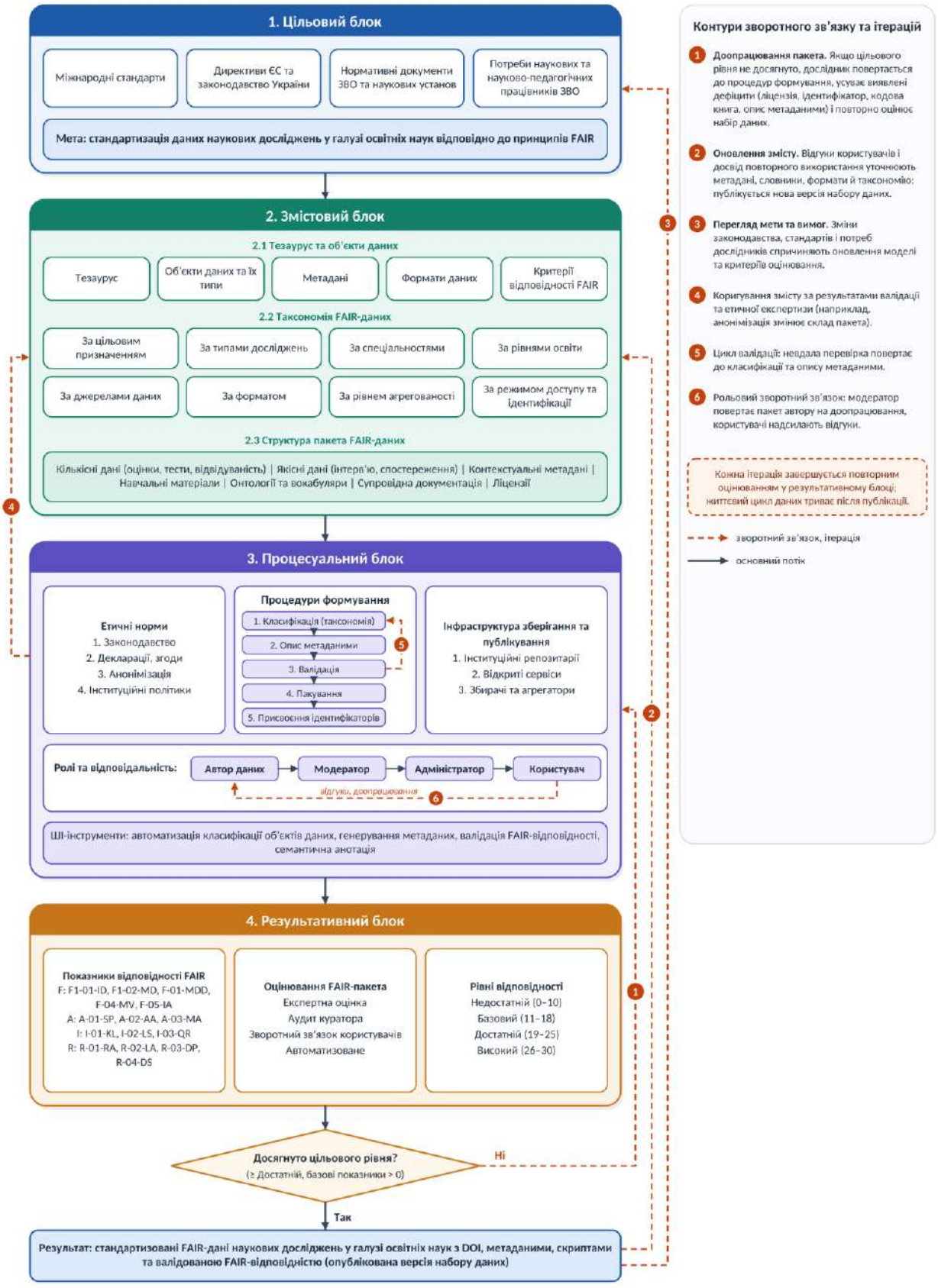


Рис. 1.1. Модель стандартизації FAIR-даних для галузі освітніх наук [8]

Змістовий блок охоплює тезаурус і типи об'єктів даних, метадані, формати, таксономію FAIR-даних (підрозділ 1.2) та структуру пакета FAIR-даних – сукупності даних, метаданих, словників, дослідницького інструментарію, дидактичних матеріалів, супровідної документації та ліцензій. Процесуальний блок описує етичні норми (законодавство, інформовані згоди, анонімізацію, інституційні політики), процедури формування пакета (класифікацію, опис метаданими, перевірку, пакування, присвоєння ідентифікаторів) та інфраструктуру зберігання й публікування (інституційні репозитарії, відкриті сервіси, агрегатори).

Результативний блок містить систему показників і рівнів, за якими оцінюють, наскільки набір даних відповідає принципам FAIR (табл. 1.1).

Таблиця 1.1.

Рівні відповідності FAIR-даних за моделлю стандартизації для галузі освітніх наук

Рівень	Діапазон балів	Обов'язкова умова	Характеристика набору даних
Недостатній (Pre-FAIR)	0–10	Базові показники оцінено нулем	Дані не готові до наукового обміну: не зберігаються взагалі або зберігаються традиційно (як додаток до статті) без ліцензії, постійного посилання та повноцінного опису
Базовий (Min-FAIR)	11–18	Щонайменше по 1 балу за кожним критерієм (F, A, I, R); немає нульових оцінок за базовими показниками	Дані опубліковано в інтернеті з URL, мають базові метадані за стандартом Дублінського ядра, текстову ліцензію й умови доступу; їх може знайти та прочитати людина
Достатній (Suff-FAIR)	19–25	Понад 50 % показників оцінено у 2 бали	Оптимальний рівень для більшості досліджень: дані збережено у структурованих форматах (CSV, JSON), описано їхній контекст і походження, присвоєно ідентифікатори
Високий (Max-FAIR)	26–30	Майже повна відповідність машинозчитуваним семантичним стандартам	Дані збережено у відкритих форматах, придатних відповідно до їх типу. Набір супроводжується повними машинозчитуваними метаданими, семантично узгодженими з стандартами та контрольованими словниками, за потреби представленими засобами RDF/JSON-LD. Забезпечено постійні ідентифікатори, документування походження, визначені умови доступу й ліцензування.

Складено за [8].

Модель не є лінійною послідовністю етапів. Її блоки пов'язані контурами зворотного зв'язку, які відображають ітеративний характер життєвого циклу даних (на рис. 1.1 їх позначено пунктиром і цифрами 1-6). Ітерації відбуваються на трьох рівнях: у межах окремих процедур, між блоками під час підготовки набору даних і після його публікації.

Внутрішні ітерації процесуального блоку забезпечують якість пакета ще до оцінювання. Якщо валідація виявляє невідповідності, пакет повертається до класифікації та опису метаданими (контур 5). Модератор так само може повернути пакет автору на доопрацювання, а користувачі можуть надсилати відгуки щодо опублікованих даних (контур 6). Результати процесу валідації можуть вимагати змінити сам склад пакета, наприклад вилучити персональні дані під час анонімізації. У такому разі процес повертається до змістового блоку (контур 4).

Основний цикл підготовки даних замикається через результативний блок. Після оцінювання за п'ятнадцятьма показниками визначають, чи досягнуто цільового рівня: щонайменше достатнього (Suff-FAIR), причому жоден базовий показник не може бути оцінений нулем. Якщо цільового рівня не досягнуто, дослідник повертається до процедур формування. Він усуває виявлені дефіцити (додає ліцензію, ідентифікатор, кодову книгу, доповнює метадані) і повторно оцінює набір даних (контур 1). Цикл повторюється, доки набір не досягне цільового рівня. Саме так результативний блок виконує функцію самоперевірки, описану вище.

Публікація не завершує життєвий цикл даних. Відгуки користувачів і досвід повторного використання можуть вимагати уточнення метаданих, словників, форматів чи таксономії. На найвищому рівні зміни законодавства, міжнародних стандартів і потреб дослідників зумовлюють перегляд мети та нормативного підґрунтя цільового блоку, а отже, оновлення самої моделі та критеріїв оцінювання (контур 3). Кожна ітерація завершується повторним оцінюванням у результативному блоці. Тому оцінку відповідності принципам FAIR слід розглядати як характеристику конкретної версії набору даних, а не як остаточний статус.

Для дослідника, який працює в галузі освіти, важливим є результативний блок: він дає змогу самостійно перевірити, наскільки підготовлений набір даних відповідає принципам FAIR, і визначити, що саме потрібно доопрацювати.

Кількісний вимір моделі становлять п’ятнадцять показників, згрупованих за чотирма критеріями F, A, I, R, кожен з яких оцінюється за трибальною шкалою від 0 до 2 балів (теоретичний максимум – 30 балів) [8].

Показники згруповано за чотирма критеріями. Відшукуваність (F): наявність унікального й постійного ідентифікатора (F1-01-ID); опис даних достатньою кількістю метаданих (F1-02-MD); зв’язок метаданих із файлами даних (F-01-MDD); машинна видимість (F-04-MV); індексація в агрегаторах (F-05-IA).

Доступність (A): доступність через стандартизований протокол (A-01-SP); наявність процедур автентифікації та авторизації (A-02-AA); збереження доступу до метаданих навіть після видалення чи закриття самих даних (A-03-MA).

Сумісність (I): використання формальної мови подання знань (I-01-KL); використання словників, що відповідають принципам FAIR (I-02-LS) з урахуванням їхнього типу; наявність кваліфікованих посилань на інші ресурси (I-03-QR).

Багаторазовість використання (R): повнота опису контексту дослідження (R-01-RA); наявність і чіткість ліцензії (R-02-LA); опис походження даних (R-03-DP); відповідність стандартам галузі освітніх наук (R-04-DS).

Сума балів за всіма показниками визначає рівень відповідності набору даних (див. табл. 1.1). Чотири показники – F1-01-ID (ідентифікатор), F1-02-MD (метадані), A-01-SP (стандартизований протокол доступу) і R-02-LA (ліцензія) – є базовими: якщо хоча б один із них оцінено нулем, набір даних не може бути віднесений навіть до базового рівня. Тому, готуючи дані до публікації, насамперед варто подбати саме про ці чотири вимоги – їх забезпечує розміщення набору даних у репозитарії з DOI, заповненим описом і зазначеною ліцензією. Наведені показники слід застосовувати так: спочатку перевіряти обов'язкові умови, потім обчислювати бали та здійснювати перехід між рівнями.

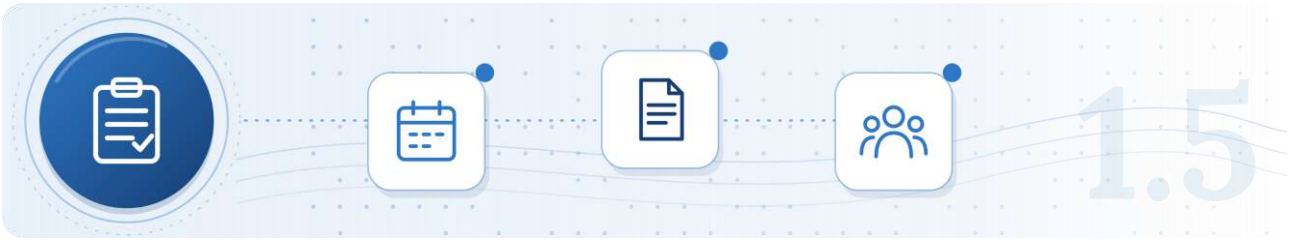
Практична цінність моделі для дослідника полягає в тому, що вона перетворює загальні принципи FAIR на перелік конкретних запитань для самоперевірки. Відповівши на них перед публікацією, можна визначити, чого бракує набору даних для досягнення достатнього рівня, наприклад, додати ліцензію, кодову книгу чи посилання на пов’язану статтю.

1.4.4. Зіставлення з міжнародними інструментами оцінювання (RDA FAIR Maturity Model, FAIRsFAIR, F-UJI)

Розроблена модель узгоджується з провідними міжнародними ініціативами оцінювання FAIR-відповідності: моделлю FAIR-зрілості Альянсу дослідницьких даних (RDA FAIR Data Maturity Model) та автоматизованим інструментом F-UJI, водночас доповнюючи їх галузевою специфікою освітніх наук, зокрема врахуванням контрольованих словників ISCED та вузькогалузевих стандартів xAPI й IEEE LOM, розглянутих у попередніх підрозділах довідника. Варто розрізняти призначення автоматизованих інструментів оцінювання: якщо F-UJI перевіряє фактичну FAIR-відповідність конкретного вже опублікованого масиву даних, то інструмент самооцінювання обізнаності FAIR-Aware, розроблений у межах проєкту FAIRsFAIR, натомість оцінює не сам масив даних, а рівень розуміння дослідником принципів FAIR, і тому є доцільним скоріше на етапі підготовки дослідника (підрозділ 2.5.1), ніж під час верифікації вже сформованого масиву (підрозділ 1.10.4). Це підтверджується й останніми міжнародними дослідженнями практики забезпечення відповідності принципам FAIR, які наголошують, що узагальнені, дисциплінарно-нейтральні критерії FAIR потребують галузевої адаптації для практичного застосування дослідниками, які не є фахівцями з управління даними [9]. Водночас емпіричні дослідження реальних наукових консорціумів засвідчують, що навіть за наявності чіткої моделі оцінювання досягнення повної FAIR-відповідності залишається ітеративним, а не одноразовим процесом, що вимагає організаційної підтримки впродовж усього життєвого циклу дослідницького проєкту [10].

Розглянуті в підрозділі 1.4 стандарти метаданих і модель стандартизації FAIR-даних для галузі освітніх наук утворюють теоретичну основу, практичне застосування якої на етапі планування дослідження розглядається в підрозділі 1.5, присвяченому планам управління даними (DMP), а на етапі вибору цифрової інфраструктури зберігання – у підрозділі 1.7.

1.5. Плани управління даними (Data Management Plans, DMP)



1.5.1. Структура і вимоги DMP

План управління даними (Data Management Plan, DMP) – це формалізований документ, що описує процедури збирання, опрацювання, документування, зберігання, захисту та оприлюднення дослідницьких даних упродовж життєвого циклу наукового дослідження й після його завершення. Складання DMP на етапі планування наукового дослідження, а не після його завершення, є визначальним чинником якості майбутнього масиву FAIR-даних, оскільки саме на цьому етапі ухвалюються рішення щодо стандарту метаданих, форматів файлів, процедур знеособлення персональних даних учасників освітнього процесу та цільового репозитарію [1]. У сучасному науковому дискурсі DMP розглядається не як адміністративна формальність, а як невіддільний складник дослідницької практики, що фіксує конкретні, верифіковні зобов'язання виконавця наукового дослідження щодо кожного з чотирьох принципів FAIR ще до початку збирання емпіричного матеріалу [1; 2].

Складання DMP є обов'язковою умовою фінансування наукових досліджень у межах рамкової програми ЄС Horizon Europe [3], а на національному рівні відповідні вимоги закріплено Методичними рекомендаціями Міністерства освіти і науки України щодо управління науковими даними для ЗВО та наукових установ [4]. Для складання DMP рекомендується використовувати спеціалізовані онлайн-інструменти – DMPonline (розробка Digital Curation Centre, Велика Британія) або Argos (сервіс OpenAIRE), які пропонують шаблони, узгоджені з вимогами конкретних організацій, що фінансують [1]. Перспективним напрямом розвитку є автоматизовані плани управління даними (machine-actionable DMP, maDMP), стандарт яких розроблено робочою групою RDA: такі плани подаються у структурованому форматі, придатному до автоматизованого опрацювання сервісами управління даними, що уможливлює автоматичне генерування частини змісту DMP і зменшує адміністративне навантаження на виконавців дослідження [2].

1.5.2. Шаблон DMP для педагогічного дослідження (заповнювана форма)

Методологічною основою наведеного нижче шаблону (табл. 1.2) слугують базові вимоги до планів управління даними, сформульовані Науковою Європою (Science Europe) у Практичному керівництві з міжнародного узгодження управління дослідницькими даними [5]. Ці вимоги розроблено спеціально для наукових установ (research performing organisations) і організацій, що фінансують, вони охоплюють шість складників, деталізованих п'ятнадцятьма орієнтувальними питаннями, і розглядаються як мінімальний стандарт, що допускає доповнення відповідно до галузевої специфіки та національного законодавства [5].

Структура шаблону узгоджена також з офіційним шаблоном DMP програми Horizon Europe [3], у якому відповідність принципам FAIR винесено в окремий структурний блок.

Шаблон адаптовано до умов виконання НДР в науковій установі: орієнтувальні питання доповнено покликаннями на відповідні підрозділи довідника, а також враховано специфіку освітніх наукових даних – соціальну чутливість емпіричного матеріалу, мультиформатність та потребу узгодження з інституційною інфраструктурою (підрозділ 1.7.4). Шаблон рекомендується заповнювати на етапі підготовки технічного завдання на НДР та актуалізувати за результатами кожного етапу її виконання.

Заповнений за наведеним шаблоном план управління даними доцільно розглядати як робочий документ, що підлягає періодичному перегляду відповідно до етапів життєвого циклу даних, розглянутих у підрозділі 1.3.1: початкова редакція, підготовлена на етапі планування НДР, уточняються за результатами збирання емпіричного матеріалу та набуває остаточного вигляду перед розміщенням масиву даних у репозитарії (підрозділ 1.10.4).

Таблиця 1.2.

Шаблон плану управління даними науково-дослідної роботи у галузі освітніх наук

№	Складник плану та орієнтовні питання	Зміст (заповнюється виконавцем)
Загальні відомості про наукове дослідження		
0.1	Повна назва науково-дослідної роботи, її державний реєстраційний номер, терміни виконання. *Напр.: «Технології формування і використання FAIR даних в*	

№	Складник плану та орієнтовні питання	Зміст (заповнюється виконавцем)
	галузі освітніх наук», ДР № 0126U000521, 2026–2027 рр.	
0.2	Наукова установа-виконавець, структурний підрозділ, науковий керівник (із зазначенням ORCID iD). *Підрозділ 1.3.2 – розподіл ролей; підрозділ 1.10.2 – ідентифікатори.*	
0.3	Версія плану, дата складання та дата останнього перегляду. *Зазначається відповідно до ітеративного характеру DMP (підрозділ 1.3.1).*	
1. Опис даних та повторне використання наявних даних		
1.1	Яким чином здійснюватиметься збирання нових та/або повторне використання наявних даних? Обґрунтувати доцільність повторного використання або причини відмови від нього. *Підрозділ 2.1 – стратегії пошуку наявних масивів даних.*	
1.2	Які типи, формати та обсяги даних передбачається отримати чи використати в межах дослідження? *Типологію подано в підрозділі 1.2; вимоги до форматів – у підрозділі 1.6.*	
2. Документування та забезпечення якості даних		
2.1	Які метадані та супровідна документація супроводжуватимуть масив даних? Який стандарт метаданих буде застосовано? *Стандарти метаданих – підрозділ 1.4; кодова книга та протокол аналізу – підрозділи 1.2.1–1.2.2.*	
2.2	Які процедури контролю якості даних буде застосовано на етапах збирання й опрацювання? *Виміри якості даних – підрозділ 2.1.2.*	
3. Зберігання та резервне копіювання під час виконання дослідження		
3.1	Яким чином забезпечуватиметься зберігання та резервне копіювання даних упродовж виконання науково-дослідної роботи?	
3.2	Яким чином забезпечуватиметься захист даних і контроль доступу до них під час виконання дослідження? *Процедури автентифікації та авторизації – підрозділ 1.8.2.*	

№	Складник плану та орієнтовні питання	Зміст (заповнюється виконавцем)
4. Правові та етичні вимоги, кодекси дослідницької поведінки		
4.1	Чи передбачає дослідження опрацювання персональних даних учасників освітнього процесу? Яким чином забезпечуватиметься дотримання вимог Закону України «Про захист персональних даних» та Загального регламенту про захист даних (GDPR)? *Нормативну базу розглянуто в підрозділі 1.9.1.*	
4.2	Які процедури знеособлення (анонімізації чи псевдонімізації) буде застосовано? Який інструментарій використовуватиметься? *Методи та інструменти – підрозділ 1.9.2.*	
4.3	Яким чином урегульовано питання майнових прав інтелектуальної власності на масив даних? *Ліцензування – підрозділ 1.8.1.*	
4.4	Чи отримано інформовану згоду учасників дослідження на оприлюднення та повторне використання даних? Які межі використання нею визначено? *Обмеження повторного використання – підрозділ 1.1.5.*	
5. Оприлюднення та довготривале збереження даних		
5.1	Яким чином і коли масив даних буде оприлюднено? Чи передбачаються обмеження доступу або період ембарго; якщо так, обґрунтувати їх. *Режими доступу – підрозділ 1.1.3.*	
5.2	У якому репозитарії буде розміщено масив даних? Обґрунтувати вибір за критеріями надійності репозитарію. *Порівняльну характеристику репозитаріїв подано в підрозділі 1.7 (табл. 1.4, 1.5).*	
5.3	Яким чином буде забезпечено застосування унікального постійного ідентифікатора (DOI, Handle) та машиночитаних метаданих? *Підрозділ 1.10.2.*	
5.4	Які масиви даних підлягають довготривалому збереженню, а які – видаленню після завершення визначеного терміну? Обґрунтувати критерії відбору. *Етап архівування – підрозділ 1.3.1.*	
5.5	Яку ліцензію буде присвоєно масиву даних?	

№	Складник плану та орієнтовні питання	Зміст (заповнюється виконавцем)
	Відкриті ліцензії – підрозділ 1.8.1.	
6. Розподіл відповідальності та ресурсне забезпечення		
6.1	Хто з виконавців науково-дослідної роботи відповідає за управління даними на кожному етапі життєвого циклу (зокрема за функції розпорядника даних)? *Ролі учасників – підрозділ 1.3.2.*	
6.2	Які ресурси (фінансові, кадрові, технічні) необхідні для реалізації плану та з яких джерел вони забезпечуються? *Витрати на знеособлення, конвертування форматів, зберігання.*	

Складено за базовими вимогами Science Europe [5] з урахуванням структури шаблону [3] та Методичних рекомендацій МОН України [4] з доповненням авторів.

Науковій установі та ЗВО доцільно оприлюднювати DMP як самостійний документ із власним постійним ідентифікатором, що відповідає практиці реєстрації планів управління даними як публічних результатів наукового дослідження та підвищує прозорість дослідницького процесу [2; 5].

1.6. Формати даних та їх зберігання



Вибір формату подання даних є технічним рішенням із довготривалими науковими наслідками: якщо масив освітніх даних збережено у форматі, підтримку якого з часом припинено, доступ до нього втрачається незалежно від того, наскільки правильно було описано метадані та наскільки надійним є репозитарій. Саме тому вимоги до форматів зберігання розглядаються в міжнародній практиці як самостійний складник управління дослідницькими даними, що виходить за межі принципів FAIR: базові вимоги Наукової Європи (Science Europe) прямо зазначають, що належне зберігання й резервне копіювання доповнюють FAIR-принципи в частині довготривалого збереження [1].

Методологічною основою організації довготривалого зберігання є еталонна модель відкритої архівної інформаційної системи OAIS (Open Archival

Information System), закріплена міжнародним стандартом ISO 14721 [2], яка визначає функційні складники архіву: приймання (ingest), архівне зберігання, управління даними, адміністрування, планування збереження, доступ і розповсюдження та регламентує міграцію цифрової інформації на нові носії й у нові формати.

1.6.1. Відкриті й пропрієтарні формати за типами освітніх даних

Основним ризиком довготривалого зберігання дослідницьких даних є технологічне старіння форматів (format obsolescence) – ситуація, за якої формат перестає підтримуватися чинним програмним забезпеченням, а також залежність від постачальника (vendor lock-in), коли відтворення даних потребує конкретного комерційного програмного продукту певної версії. Для освітніх наукових даних ці ризики є особливо відчутними, оскільки значна частина емпіричного матеріалу традиційно накопичується у пропрієтарних форматах статистичних пакетів (SPSS, Statistica) та офісних застосунків, а якісні дані у форматах, що залежать від конкретного програмного засобу якісного аналізу (NVivo, MAXQDA, ATLAS.ti).

У міжнародній практиці цифрового збереження сформульовано низку критеріїв, за якими формат визнається придатним для довготривалого зберігання [3]: відкритість і публічна доступність специфікації формату; незалежність від конкретного постачальника програмного забезпечення чи операційної системи; поширеність і підтримка широким колом програмних засобів; відсутність стиснення з втратами; відсутність технічних засобів захисту, що обмежують доступ (шифрування, механізми управління цифровими правами); самоописовість – здатність формату містити структурні метадані, що уможливлюють інтерпретацію вмісту без зовнішньої документації. Відповідність цим критеріям безпосередньо корелює з рівнем інтероперабельності масиву даних, розглянутим у підрозділі 1.1.4: формати, що кодують семантику даних (RDF, JSON-LD), одночасно задовольняють вимоги інтероперабельності та вимоги придатності до тривалого зберігання.

Практичною стратегією, рекомендованою для наукових установ і ЗВО, є нормалізація форматів (format normalisation) – конвертування даних у відкритий формат зберігання з одночасним збереженням первинного файлу в оригінальному форматі. Такий підхід дає змогу, з одного боку, гарантувати довготривалу доступність масиву даних, а з іншого – зберегти повноту первинної інформації, оскільки конвертування може супроводжуватися втратою окремих елементів (форматування змінних, міток значень, вбудованих

формул). Обидві версії масиву, первинну й нормалізовану, доцільно розміщувати в репозитарії як складники єдиного пакета даних із відповідним описом у метаданих, що узгоджується з вимогою документування походження даних (provenance) (підрозділ 1.1.5).

Довготривале збереження передбачає також періодичний контроль цілісності файлів за допомогою контрольних сум (checksums) – обчислених за алгоритмами MD5 чи SHA-256 значень, що дають змогу виявити пошкодження бітового потоку під час зберігання чи передавання, та планову міграцію масивів даних у нові формати в разі старіння наявних. У надійних репозитаріях, розглянутих у підрозділі 1.7, ці процедури виконуються автоматично на рівні інфраструктури, що є одним із визначальних критеріїв вибору репозитарію для розміщення результатів НДР.

1.6.2. Порівняльна характеристика форматів за придатністю до довготривалого зберігання

У таблиці 1.3 подано узагальнену характеристику форматів подання освітніх наукових даних, згрупованих за типами даних відповідно до типології (підрозділ 1.2). Рекомендації сформульовано з урахуванням Переліку рекомендованих форматів Бібліотеки Конгресу США [3], настанов Служби даних Великої Британії [4] та базових вимог Наукової Європи [1].

Таблиця 1.3.

Узагальнена характеристика форматів подання освітніх наукових даних

Тип даних	Робочі формати	Рекомендовані формати зберігання	Примітки
Табличні кількісні дані (результати тестувань, анкетування)	XLSX, SAV (SPSS), DTA (Stata), RDATA	CSV, TSV (із супровідною кодовою книгою)	Конвертування у CSV спричиняє втрату міток змінних і значень, тому кодова книга є обов'язковою (підрозділ 1.2.1)
Напівструктуровані дані (журнали активності, xAPI)	Власні формати систем управління навчанням, LOG	JSON, XML, JSON-LD	Самоописові формати; JSON-LD додатково забезпечує семантичну сумісність (підрозділ 1.1.4)
Текстові та якісні дані (транскрипти інтерв'ю, польові нотатки)	DOCX, ODT, власні формати NVivo, MAXQDA, ATLAS.ti	TXT (UTF-8), RTF, PDF/A (ISO 19005)	Кодову схему якісного аналізу експортувати окремо у відкритому форматі
Аудіозаписи	MP3, WMA, M4A	WAV (PCM), FLAC	Формати без втрат; MP3

Тип даних	Робочі формати	Рекомендовані формати зберігання	Примітки
(інтерв'ю, фокус-групи)			припустимий лише як версія для ознайомлення
Відеозаписи навчальних занять	AVI, MOV, WMV	MP4 (H.264/AAC), MKV (FFV1)	Зважати на потребу знеособлення зображень учасників (підрозділ 1.9.2)
Зображення (сканкопії робіт, фотоматеріали)	JPEG, BMP, PSD	TIFF (без стиснення), PNG	JPEG застосовує стиснення з втратами й не рекомендований як формат зберігання
Онтології та контрольовані словники	Власні формати редакторів онтологій	RDF/XML, Turtle, OWL, SKOS	Забезпечують найвищий рівень інтероперабельності (підрозділ 1.1.4)
Геопросторові дані (розташування закладів освіти)	SHP (ESRI Shapefile), власні формати ГІС	GeoJSON, GML, GeoTIFF	Обов'язково зазначати систему координат у метаданих
Програмний код опрацювання даних (скрипти аналізу)	Записники Jupyter, проєктні файли середовищ розробки	Вихідний код у текстовому форматі (R, PY) із файлом залежностей	Належить до дослідницького програмного забезпечення; застосовуються принципи FAIR4RS

Складено за [3; 4] з урахуванням типології освітніх наукових даних (підрозділ 1.2) і з доповненням авторів.

Наведені рекомендації доцільно фіксувати в плані управління даними НДР (підрозділ 1.5.2, питання 1.2 та 5.4) ще на етапі планування дослідження, оскільки вибір формату збирання первинних даних значною мірою визначає обсяг подальших витрат на конвертування. Практика засвідчує, що узгодження форматів на початку виконання НДР є істотно менш витратним, ніж нормалізація вже накопиченого різнорідного масиву даних перед його розміщенням у репозитарії.

1.7. Цифрові сервіси та репозитарії для зберігання FAIR-даних



Реалізація принципів FAIR, розглянутих у підрозділі 1.1, потребує конкретної цифрової інфраструктури (репозитаріїв, спеціалізованих сховищ і

хмарних платформ), що здатна забезпечити довготривале зберігання, відшукуваність і контрольований доступ до освітніх наукових даних. Вибір репозитарію є рішенням, що безпосередньо визначає рівень FAIR-відповідності опублікованого масиву даних, оскільки саме інфраструктура забезпечує присвоєння постійних ідентифікаторів, індексацію метаданих у глобальних агрегаторах, контроль цілісності файлів і міграцію форматів (підрозділ 1.6). У цьому підрозділі послідовно розглянуто шість рівнів такої інфраструктури: загальнонаукові репозитарії, спеціалізовані освітні сховища, програмні платформи інституційних репозитаріїв, національну та інституційну інфраструктуру України, захищені архіви для чутливих даних, а також Європейську хмару відкритої науки.

1.7.1. Загальнонаукові репозитарії даних

Репозитарій даних – це цифрова інфраструктура для довготривалого зберігання, опису та поширення наукових масивів даних. У межах дослідження проаналізовано чотири провідні загальнонаукові репозитарії, придатні для оприлюднення освітніх наукових даних: Zenodo, Figshare, OSF (Open Science Framework) та Dryad (табл. 1.4) [1].

Zenodo (zenodo.org) – безкоштовний мультидисциплінарний репозитарій, розроблений Європейською організацією з ядерних досліджень (CERN) за підтримки програми «Горизонт 2020», рекомендований Європейською комісією та OpenAIRE для оприлюднення результатів проєктів Horizon Europe. Підтримує будь-які формати файлів, автоматично присвоює DOI, забезпечує повне версіонування з об’єднувальним «концептуальним» ідентифікатором; обмеження обсягу 50 ГБ на масив даних. Режим доступу налаштовується диференційовано (відкритий, обмежений, закритий з відкритими метаданими, ембарго).

Figshare (figshare.com) – платформа, зручна для оприлюднення рисунків, таблиць, презентацій і плакатів завдяки вбудованому попередньому перегляду файлів; передбачає безкоштовний тариф для невеликих масивів і платні інституційні тарифи для необмеженого обсягу, а також широку інтеграцію з науковими видавцями.

OSF (osf.io) – платформа підтримки повного циклу дослідницького проєкту з можливістю попередньої реєстрації дослідження (preregistration), ієрархічною структурою проєкту й безкоштовним доступом; обмеження – 5 ГБ на файл.

Dryad (datadryad.org) – репозитарій з обов'язковим кураторським опрацюванням: кожен масив даних перевіряється куратором на повноту метаданих; обсяг – до 300 ГБ, вартість депонування – 120 доларів США за відсутності інституційного членства, наявна інтеграція з ORCID та провідними науковими видавцями [1].

Основні характеристики розглянутих репозитаріїв, зокрема умови використання, присвоєння DOI, обмеження обсягу та підтримку галузевих стандартів метаданих, порівняно в табл. 1.4.

Таблиця 1.4.

Порівняльна характеристика загальнонаукових репозитаріїв

Характеристика	Zenodo	Figshare	OSF	Dryad
Умови використання	Безкоштовно	Безкоштовно / платно	Безкоштовно	від 150 USD (масив до 5 ГБ); для партнерів Dryad – за рахунок установи
Присвоєння DOI	Автоматично	Автоматично	За запитом (для публічних проєктів); для реєстрацій – автоматично	Автоматично
Версіонування	Повне	Так	Так	Обмежене
Кураторський контроль	Немає	Немає	Немає	Обов'язковий
Попередня реєстрація дослідження	Немає	Немає	Так	Немає
Обмеження обсягу	50 ГБ	Необмежено	5 ГБ на файл	300 ГБ
Статус у Horizon Europe / EOSC	Рекомендований	Частковий	Частковий	Частковий
Підтримка стандарту DDI	Немає	Немає	Немає	Немає

Складено за [1].

Спільним обмеженням усіх чотирьох сервісів є відсутність нативної підтримки стандарту DDI, розглянутого в підрозділах 1.1.4 та 1.2.1, що є особливо істотним для структурованих педагогічних досліджень, які ґрунтуються на анкетуванні й тестуванні [1]. Це зумовлює доцільність звернення до спеціалізованих освітніх сховищ там, де це можливо.

1.7.2. Спеціалізовані освітні сховища та відкриті бази даних

На відміну від загальнонаукових репозитаріїв, спеціалізовані сховища орієнтовані на конкретний тип освітніх даних і, як правило, підтримують галузеві стандарти метаданих (підрозділ 1.4). ICPSR (Inter-university Consortium for Political and Social Research) – провідний архів даних соціальних і поведінкових наук зі значним масивом освітніх даних, що підтримує стандарт DDI, забезпечує довготривале кураторське опрацювання та надає доступ дослідникам установ-членів консорціуму [2]. Міжнародна асоціація з оцінювання освітніх досягнень (IEA) підтримує міжнародну базу даних TIMSS & PIRLS, яка надає у відкритому доступі первинні мікродані міжнародних порівняльних досліджень із математики, природничих наук і читацької грамотності у форматах SPSS та SAS разом із докладними посібниками користувача [3]. Аналогічну функцію для програми міжнародного оцінювання учнів PISA виконує портал відкритих даних Організації економічного співробітництва та розвитку [4].

Окрему категорію становлять реєстри й агрегатори, які не зберігають дані безпосередньо, а забезпечують пошук профільного репозитарію: re3data.org – міжнародний реєстр понад 3000 репозитаріїв даних із фільтрацією за предметною галуззю, типом доступу й підтримуваними стандартами [5]; FAIRsharing.org – реєстр стандартів метаданих, баз даних і політик управління даними з функцією перехресного пошуку [6]. Для відкритих освітніх ресурсів (навчальних матеріалів, а не власне дослідницьких даних) подібну функцію виконують міжнародні платформи MERLOT та OER Commons, які, хоча й не є репозитаріями наукових даних у строгому розумінні, часто слугують джерелом артефактів освітньої діяльності, розглянутих у типології підрозділу 1.2.

1.7.3. Платформи інституційних репозитаріїв і бібліотечних систем: порівняння за критеріями FAIR

Розглянуті вище сервіси є готовими до використання платформами, тоді як наукова установа чи ЗВО, що розгортають власний репозитарій, постають перед вибором програмної платформи.

Для узагальнення результатів аналізу цифрових платформ проведено порівняльний аналіз семи цифрових платформ – інституційних репозитаріїв (DSpace, EPrints, Invenio RDM, Fedora/Samvera, Greenstone) та бібліотечних систем (Koha, Ex Libris Alma/Primo) щодо відповідності основним принципам FAIR. Джерелами для оцінювання слугували функціональні можливості платформ та їхня офіційна документація. Як інструмент використано

адаптовану матрицю критеріїв (F1–F4, A1–A2, I1–I3, R1–R3), побудовану на основі метрик F-UJI. Щоб результати можна було зіставити, застосовано трирівневу шкалу: 3 бали відповідають високому рівню відповідності, 2 бали – середньому, 1 бал – низькому (табл. 1.5) [7].

Таблиця 1.5.

Відповідність платформ інституційних репозитаріїв і бібліотечних систем принципам FAIR

Платформа	Оцінка, балів (F / A / I / R	Характеристика відповідності
Invenio RDM	12/12 (3 / 3 / 3 / 3)	Найвищий рівень відповідності: вбудована підтримка DOI, схеми метаданих DataCite, машиночитаних ліцензій, узгодженість із настановами OpenAIRE та EOSC
Fedora/ Samvera	11/12 (2 / 3 / 3 / 3)	Гнучка семантична архітектура, підтримка зв'язаних даних; потребує значних технічних і кадрових ресурсів для впровадження
Ex Libris Alma/Primo	11/12 (3 / 3 / 3 / 2)	Підтримка BIBFRAME, Schema.org і DataCite; повноцінна реалізація принципу Reusable забезпечується не повністю
DSpace	8/12 (2 / 2 / 2 / 2)	Середній рівень: підтримка Handle та протоколу OAI-PMH, однак обмежена реалізація принципів Interoperable і Reusable
EPrints	8/12 (2 / 2 / 2 / 2)	Середній рівень із характеристиками, подібними до DSpace
Koha	7/12 (2 / 2 / 2 / 1)	Часткова відповідність, забезпечувана переважно через формат MARC 21 і протокол Z39.50
Greenstone	4/12 (1 / 1 / 1 / 1)	Мінімальний рівень відповідності; орієнтована на офлайн-середовища

Складено за [7] із доповненням авторів.

За результатами аналізу для закладів освіти й наукових установ України рекомендовано модернізацію наявних репозитаріїв на платформі DSpace через інтеграцію зі службою DataCite, а в перспективі – міграцію на платформу Invenio RDM у контексті розбудови FAIR-інфраструктури освітніх наук [7].

1.7.4. Національна та інституційна інфраструктура

В Україні функцію центрального вузла інституційного накопичення наукових результатів виконує Національний репозитарій академічних текстів (НРАТ, nrat.ukrintei.ua) – загальнодержавна розподілена електронна база даних, власником якої є держава в особі Міністерства освіти і науки України, а

розпорядником – Український інститут науково-технічної експертизи та інформації (УкрІНТЕІ). НРАТ побудовано за федеративним принципом: він складається з центрального репозитарію та мережі локальних інституційних репозитаріїв ЗВО й наукових установ, підключення яких технічно підтримується для найпоширеніших платформ – DSpace, EPrints, OJS. НРАТ зареєстровано як джерело бібліотечних і наукових даних у міжнародних реєстрах OpenDOAR і FAIRsharing, що безпосередньо реалізує складник F4 принципу Findable, розглянутий у підрозділі 1.1.2 [8].

Електронна бібліотека Національної академії педагогічних наук України (НАПН України) (*lib.iitta.gov.ua*) надає відкритий доступ до науково-освітніх інформаційних ресурсів установ НАПН. Технічну підтримку й адміністрування бібліотеки здійснює Інститут цифровізації освіти НАПН України (ІЦО НАПН України) [9]. Бібліотека функціонує на платформі EPrints і містить станом на жовтень 2026 року понад 42 тисячі ресурсів: наукові публікації, дисертації, матеріали конференцій, аудіо- та відеозаписи семінарів. Вона є першим кроком для розміщення метаданих і публікацій, пов'язаних із результатами педагогічних досліджень, перш ніж власне дослідницькі дані буде оприлюднено у спеціалізованому репозитарії даних (підрозділ 1.7.1). З огляду на результати порівняльного аналізу платформ (підрозділ 1.7.3) перспективним напрямом розвитку інституційної інфраструктури НАПН України є інтеграція наявної платформи зі службою DataCite для автоматичного присвоєння DOI масивам дослідницьких даних.

1.7.5. Захищені інституційні архіви для чутливих освітніх даних: досвід FAIRVault

Специфіка освітніх наукових даних, розглянута в підрозділі 1.2, полягає в їх соціальній чутливості: значна частина емпіричного матеріалу стосується учасників освітнього процесу, зокрема неповнолітніх, і не може бути розміщена у відкритих загальнонаукових репозитаріях без порушення вимог законодавства про захист персональних даних (підрозділ 1.9). Відповіддю на цю потребу є окремий клас інфраструктурних рішень – захищені архіви з контрольованим доступом, показовим прикладом яких є міжуніверситетський сервіс FAIRVault, створений чотирма фламандськими університетами (Ghent University, Hasselt University, University of Antwerp, Vrije Universiteit Brussel) для безпечного довготривалого зберігання, архівування та контрольованого доступу до наукових даних, які недоцільно або неможливо розміщувати в наявних зовнішніх репозитаріях [10].

Архітектурно FAIRVault поєднує дві функції: пакувальний механізм із федеративним рішенням для зберігання, за якого пакети даних зберігаються в

локальних сховищах кожної установи-партнера, та портал доступу до метаданих, через який дослідники можуть шукати масиви даних і завантажувати їх або надсилати запит на доступ залежно від умов, визначених депонентом [11]. Ключовими функційними можливостями сервісу є диференціювання рівнів чутливості даних за допомогою позначок (Data Tags), запровадження комітетів з доступу до даних (Data Access Committee), інтеграція зі службами DataCite та FRIS, а також принцип незмінної відкритості метаданих: опис масиву даних залишається публічно доступним навіть після вилучення самих даних, що є прямою реалізацією складника A2 принципу Accessible (підрозділ 1.1.3).

Досвід FAIRVault демонструє практичну реалізацію засадничого принципу відкритої науки – «настільки відкрито, наскільки можливо, настільки закрито, наскільки необхідно» – стосовно саме тієї категорії даних, яка переважає в педагогічних дослідженнях. Запровадження подібного міжінституційного рішення на національному рівні дало б змогу оприлюднювати метадані чутливих освітніх масивів даних, забезпечуючи їх відшукуваність для наукової спільноти без порушення прав учасників освітнього процесу.

1.7.6. Європейська хмара відкритої науки (EOSC) та її сервіси для українських дослідників

Європейська хмара відкритої науки (European Open Science Cloud, EOSC) – федеративна інфраструктура, що об’єднує наукові сховища, сервіси та масиви даних країн ЄС й асоційованих держав у єдиному багатодисциплінарному середовищі для оприлюднення, пошуку та повторного використання результатів досліджень [12]. Управління EOSC здійснюється за трипартитною моделлю за участю Європейської комісії, Керівної ради EOSC та Асоціації EOSC; функціонує єдиний вузол EOSC EU Node, що надає дослідникам уніфікований доступ до інструментів, обчислювальних ресурсів і сервісів зберігання даних у межах федерації [13].

Для українських дослідників участь у просторі EOSC ґрунтується на угоді щодо асоційованої участі України в програмі Horizon Europe, підписаній у жовтні 2021 р.; з жовтня 2023 р. на базі Національного фонду досліджень України функціонує офіс Horizon Europe в Україні. Інтеграція України в європейський дослідницький простір регулюється також Дорожньою картою інтеграції наукової та інноваційної системи України в європейський дослідницький простір та Концепцією Державної цільової програми розвитку дослідницьких інфраструктур [14]. НРАТ, розглянутий у підрозділі 1.7.4, є практичним

національним будівельним блоком, вже сумісним із принципами й реєстрами EOSC, і в перспективі може набути статусу українського вузла федерації.

Практичне значення участі в EOSC для дослідників галузі освітніх наук полягає в кількох аспектах: доступ до єдиного пошукового інтерфейсу за масивами даних і публікаціями з усієї Європи, що розширює можливості пошуку споріднених досліджень і партнерів для міжнародних порівняльних досліджень; доступ до спільних обчислювальних ресурсів для опрацювання великих масивів освітніх даних; підвищення видимості та цитованості власних наукових результатів завдяки оприлюдненню в репозитаріях, офіційно інтегрованих з EOSC, насамперед у Zenodo; відповідність вимогам грантових заявок у межах Horizon Europe, які дедалі частіше передбачають оприлюднення дослідницьких даних у форматі, сумісному з інфраструктурою EOSC. Розглянута багаторівнева інфраструктура створює технічну основу для практичних процедур оприлюднення масивів даних (підрозділ 1.10).

1.8. Доступ, ліцензування та права



1.8.1. Ліцензування: Creative Commons та Open Data Commons (ODC)

Ліцензування є юридичним інструментом, що унормовує умови фактичного використання масиву даних після отримання доступу до нього, і безпосередньо реалізує складник R1.1 принципу Reusable (підрозділ 1.1.5). Без явно зазначеної, машиночитаної ліцензії дані за замовчуванням залишаються під дією повного авторського права, що унеможливлює їх використання третьою стороною навіть за наявності технічного доступу.

Для більшості типів освітніх наукових даних (якісних матеріалів, звітів, окремих файлів) застосовними є ліцензії Creative Commons (CC), розглянуті в контексті принципу Reusable: Attribution 4.0 International (CC BY 4.0), що дозволяє будь-яке використання за умови зазначення авторства, та CC0 1.0 Universal, що передає матеріал у суспільне надбання [1].

Водночас структуровані табличні масиви даних, які законодавство про авторське право і суміжні права розглядає як бази даних, в окремих юрисдикціях, зокрема в ЄС, охороняються не лише авторським правом на їхній вміст, а й

особливим правом виробника бази даних (Директива 96/9/ЄС). Це право прямо охоплюють лише ліцензії Creative Commons версії 4.0; попередні версії його не врегульовували. Тому для оприлюднення наборів даних слід обирати ліцензії CC 4.0 або ліцензії Open Data Commons (ODC), які Open Knowledge Foundation розробила спеціально для баз даних. Таких ліцензій три. Ліцензія із зазначенням авторства (ODC Attribution License, ODC-BY) дозволяє будь-яке повторне використання бази даних за умови зазначення її автора. Відкрита ліцензія для баз даних (Open Database License, ODbL) до вимоги зазначати авторство додає принцип «копілефту» (share-alike): похідну базу даних потрібно поширювати на умовах тієї самої ліцензії. Ліцензія на передання в суспільне надбання (Public Domain Dedication and License, PDDL) є аналогом CC0 для баз даних: вона передає базу даних у суспільне надбання без будь-яких обмежень [2].

Основні характеристики розглянутих відкритих ліцензій, зокрема умови зазначення авторства, допустимість комерційного використання та створення похідних наборів даних, узагальнено у табл. 1.6.

Таблиця 1.6.

Порівняння відкритих ліцензій

Ліцензія	Дозволене використання	Вимоги	Рекомендований тип даних
CC BY 4.0	Будь-яке, зокрема комерційне, і створення похідних матеріалів	Зазначення авторства	Набори даних, публікації, звіти, якісні матеріали, окремі файли
CC0 1.0	Необмежене (передання у суспільне надбання)	Немає (посилання на джерело – за нормами академічної доброчесності)	Агреговані статистичні показники, метадані
ODC-BY	Будь-яке використання бази даних, зокрема комерційне	Зазначення авторства	Структуровані табличні масиви і бази даних
ODbL	Будь-яке, зокрема комерційне, і створення похідних баз даних	Зазначення авторства; поширення похідних баз даних на тих самих умовах	Масиви даних, які спільно розробляють чи постійно оновлюють
ODC PDDL	Необмежене (передання в суспільне надбання)	Немає (посилання на джерело – за нормами академічної доброчесності)	Відкриті адміністративні дані

Складено за [2; 3]. з доповненням авторів.

Незалежно від обраної ліцензії принципове значення має її розміщення безпосередньо в структурованих метаданих масиву даних (полі license схеми

DataCite), а не лише текстом у файлі README чи на сторінці репозитарію, – це надає можливість автоматизованої перевірки умов використання програмними агентами й агрегаторами (підрозділ 1.7.2).

1.8.2. Автентифікація й авторизація доступу

Автентифікація (перевірка особи користувача) й авторизація (надання йому прав доступу) забезпечують контрольований доступ до даних обмеженого поширення, при цьому метадані таких даних залишаються відкритими й доступними для пошуку (підрозділ 1.1.2). У цифрових репозитаріях розрізняють автоматизовані процедури – вхід через ORCID iD або через інституційний обліковий запис установи (єдиний вхід, SSO) та ручні процедури: запит доступу електронною поштою до розпорядника даних або розгляд заявки комітетом з доступу до даних (Data Access Committee), як у сервісі FAIRVault (підрозділ 1.7.5). Автоматизовані процедури швидші й зручніші, тоді як ручні, попри більше навантаження, залишаються необхідними для найчутливіших даних, зокрема даних про неповнолітніх (підрозділ 1.9).

Для дослідника важливо розрізняти три поширені режими доступу, які пропонують репозитарії. За відкритого доступу дані може завантажити будь-хто без реєстрації, а автентифікація потрібна лише авторові – для створення й редагування запису. За обмеженого доступу опис набору даних відкритий, але файли надаються лише після запиту: користувач входить до системи (найчастіше через ORCID iD або обліковий запис установи), пояснює мету використання, а автор чи розпорядник даних ухвалює рішення. Такий режим, зокрема, підтримує Zenodo, де автор сам розглядає запити й може визначити умови, на яких надає доступ [22]. За закритого доступу файли не надаються зовсім, однак метадані залишаються відкритими – це дає змогу іншим дослідникам дізнатися про існування набору даних і за потреби звернутися до автора.

1.9. Етичні та правові аспекти захисту персональних даних



1.9.1. Нормативна база: GDPR і Закон України «Про захист персональних даних»

Специфіка освітніх наукових даних, розглянута в підрозділі 1.2, полягає в тому, що значна їх частина прямо чи опосередковано стосується конкретних

учасників освітнього процесу – здобувачів освіти, зокрема неповнолітніх, педагогічних працівників, батьків, – і тому підпадає під дію законодавства про захист персональних даних. В Україні відповідні відносини регулює Закон України «Про захист персональних даних» від 01.06.2010 № 2297-VI, який визначає персональні дані як будь-яку інформацію, що прямо чи опосередковано стосується фізичної особи, яку ідентифіковано або може бути ідентифіковано [1]. У Європейському Союзі аналогічну функцію виконує Загальний регламент про захист даних (GDPR, Регламент (ЄС) 2016/679), що застосовується до опрацювання персональних даних осіб, які перебувають в ЄС, а також до опрацювання даних установами, що діють в ЄС (ст. 3 GDPR), і слугує орієнтиром для гармонізації вітчизняного законодавства в межах євроінтеграційних процесів [2].

Правовою підставою опрацювання персональних даних учасників освітнього процесу в межах наукового дослідження, як правило, є інформована згода – добровільне, конкретне, поінформоване й однозначне волевиявлення суб'єкта персональних даних. Для дослідника принципове значення має розмежування трьох станів даних: персональні дані (містять прямі чи непрямі ідентифікатори), псевдонімізовані дані (ідентифікатори замінено на умовний код, а ключ відповідності зберігається окремо, тому теоретична можливість повторної ідентифікації залишається) та анонімізовані дані (ідентифікація неможлива жодним розумно доступним способом, тому такі дані виходять за межі дії законодавства про захист персональних даних) [2]. Саме останній стан є цільовим для масивів, що плануються до відкритого оприлюднення в репозитаріях, розглянутих у підрозділі 1.7, тоді як псевдонімізовані дані можуть поширюватися лише в режимі обмеженого доступу з процедурою автентифікації (підрозділ 1.8.2).

1.9.2. Анонімізація та псевдонімізація: методи й інструменти (ARX, Amnesia, sdcMicro, OpenDP)

Анонімізація освітніх даних на практиці передбачає кілька послідовних кроків. Насамперед із масиву вилучають прямі ідентифікатори – прізвища та імена, адреси електронної пошти, номери телефонів, номери учнівських чи студентських квитків. Їх замінюють умовними кодами (псевдонімізація), а таблицю відповідності кодів і осіб зберігають окремо з обмеженим доступом або знищують, якщо вона більше не потрібна. Далі звертають увагу на непрямі ідентифікатори (квазіідентифікатори) – вік, стать, клас, заклад освіти, населений пункт, тому що окремо вони не вказують на особу, але їх поєднання

в невеликій вибірці може це зробити (наприклад, «єдина вчителька інформатики віком 30–35 років у сільській школі району»). Такі значення узагальнюють, тобто замінюють точний вік віковою групою, назву закладу – його типом – або вилучають рідкісні значення. Окремо перевіряють відкриті відповіді анкет, транскрипти інтерв'ю та відеозаписи, у яких учасники можуть згадувати імена чи інші впізнавані деталі. Програмні інструменти анонімізації здебільшого спираються на модель k-анонімності: дані перетворюють так, щоб кожен запис за непрямими ідентифікаторами збігався щонайменше з k−1 іншими, тобто кожна особа «губилася» в групі щонайменше з k осіб. Для найчутливіших даних застосовують також математично строгіший підхід – диференційну приватність, за якого до результатів статистичних обчислень додають контрольовані випадкові відхилення, щоб за оприлюдненими статистичними показниками не можна було дізнатися відомості про конкретну особу.

Для дослідників, які не є фахівцями з інформаційної безпеки, найзручнішими є інструменти з графічним інтерфейсом. ARX – безкоштовна програма анонімізації, яка, крім власне перетворення даних, показує, наскільки зменшилася їхня інформативність [3]. Amnesia – інструмент інфраструктури OpenAIRE, що працює як вебсервіс або локально, орієнтований на просту візуальну роботу та підготовку даних до публікації у відкритих репозитаріях [4]. Пакет sdcMicro для мови програмування R [5] та бібліотека OpenDP, розвинута Гарвардським університетом спільно з компанією Microsoft [6], потребують навичок програмування; їх доцільно застосовувати разом із фахівцями зі статистики чи IT-супроводу (підрозділ 1.3.2).

Основні можливості найпоширеніших безкоштовних інструментів анонімізації, зокрема методи, які вони реалізують, тип інтерфейсу та умови, за яких кожен з них доцільно застосовувати, порівняно в табл. 1.7.

Таблиця 1.7.

Порівняльна характеристика інструментів анонімізації

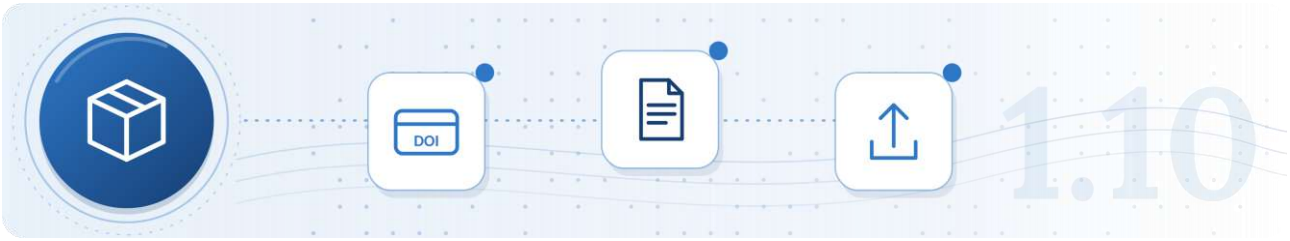
Інструмент	Методи	Інтерфейс	Найкраще застосування
ARX	k-анонімність, l-різноманітність, t-близькість, диференційна приватність, оцінювання ризику повторної ідентифікації	Графічний, кросплатформний (Windows, macOS, Linux)	Складні багатовимірні табличні масиви (наприклад, результати анкетування з багатьма демографічними ознаками)
Amnesia	k-анонімність, km-анонімність	Графічний; локальна	Швидка анонімізація невеликих табличних

Інструмент	Методи	Інтерфейс	Найкраще застосування
		програма або онлайн-версія	масивів перед публікацією у відкритих репозитаріях
sdcMicro	Узагальнення (перекодування), приховування окремих значень, мікроагрегація, внесення випадкових відхилень, оцінювання ризику повторної ідентифікації	Пакет для R (скрипти) або графічний вебінтерфейс sdcApp	Дослідники з навичками роботи в R; масиви офіційної статистики й великі опитування
OpenDP	Диференційна приватність	Програмна бібліотека (Python; основа – Rust)	Статистичні запити до чутливих масивів без оприлюднення первинних даних (наприклад, середні бали за закладами освіти)

Складено за [4; 5; 7; 8] із доповненням авторів.

Вибір конкретного інструменту й методу анонімізації доцільно фіксувати в плані управління даними НДР (підрозділ 1.5.2, питання 4.2) ще на етапі планування дослідження, оскільки надмірне узагальнення чи приховування окремих значень знижує аналітичну цінність масиву даних, а недостатнє – залишає ризик повторної ідентифікації учасників освітнього процесу. Досягнення балансу між захистом приватності й корисністю даних (privacy-utility trade-off) є завданням, що потребує узгодженого рішення керівника дослідження, розпорядника даних і фахівця із захисту персональних даних, ролі яких розглянуто в підрозділі 1.3.2.

1.10. Пакування, ідентифікація та публікація даних



Цей підрозділ завершує виклад практичного циклу формування FAIR-даних, розпочатий у підрозділі 1.3.1: якщо попередні підрозділи розглядали окремі складники – стандартизацію (1.4), планування (1.5), формати (1.6), інфраструктуру (1.7), ліцензування (1.8) та захист персональних даних (1.9), то у даному підпункті ці складники об’єднуються в послідовність дій із пакування, ідентифікації та безпосереднього оприлюднення масиву даних у репозитарії.

1.10.1. Стандарти пакування дослідницьких об’єктів (RO-Crate, BagIt)

Дослідницький масив даних рідко складається з одного файлу: типовий пакет освітніх FAIR-даних, розглянутий у підрозділі 1.4.3, охоплює первинні дані, кодову книгу чи протокол аналізу (підрозділи 1.2.1–1.2.2), скрипти опрацювання (підрозділ 2.3.2) та супровідні метадані. Стандарти пакування визначають, яким чином структурувати сукупність цих різнорідних файлів в один логічно цілісний, самоописовий і машиночитаний об’єкт, придатний до передавання й довготривалого зберігання.

На практиці пакет даних – це впорядкована папка з файлами, яка містить щонайменше: первинні дані у відкритому форматі; кодову книгу чи протокол аналізу; файл README з поясненням змісту, структури, методики збирання й умов використання даних; файл ліцензії; за потреби – анкети, тести та інший інструментарій, а також скрипти опрацювання. Файли доцільно називати зрозуміло й однаково (наприклад, «2026_opytuvannia_vchyteliv_v1.csv»), а в README пояснити прийняті позначення. Для машиночитаного опису такого пакета розроблено стандарт RO-Crate (Research Object Crate): до папки додається один файл з описом кожного елемента та зв’язків між ними – хто створив дані, яким інструментом, на яких умовах їх можна використовувати [9].

Інший стандарт – BagIt [1] зосереджений на цілісності пакета: до папки з даними додається перелік усіх файлів із контрольними сумами, що дає змогу перевірити, чи не змінилися файли під час передавання чи тривалого зберігання. Обидва стандарти доповнюють один одного; як правило, їх застосовують репозитарії та архіви цифрового збереження, зокрема Бібліотека Конгресу США та репозитарій Dryad. Завдання дослідника – підготувати впорядковану структуру пакета й інформативний файл README, що вже суттєво підвищує придатність даних до повторного використання.

1.10.2. Присвоєння постійних ідентифікаторів (DOI, ORCID, ROR)

Постійний ідентифікатор (Persistent Identifier, PID) – це незмінне «цифрове посилання», за яким набір даних, автора чи установу можна однозначно знайти навіть через багато років, коли звичайні вебадреси вже не працюють (підрозділ 1.1.2). Досліднику-освітянину достатньо подбати про три такі ідентифікатори.

DOI набору даних. Окремо його отримувати не потрібно: репозитарій (Zenodo, Figshare, OSF, Dryad) присвоює DOI автоматично під час публікації набору даних (підрозділ 1.10.4). Саме DOI варто зазначати у статті, дисертації чи звіті, коли ви посилаєтеся на власні дані.

ORCID iD автора. Це особистий номер науковця, що пов'язує з ним усі його публікації й набори даних, навіть якщо змінилося прізвище чи місце роботи. Реєстрація на сайті orcid.org безкоштовна і триває кілька хвилин; ORCID iD варто зазначати під час кожного завантаження даних до репозитарію.

ROR ID установи (Research Organization Registry (реєстр ідентифікаторів наукових організацій)). Це ідентифікатор організації, у якій виконано дослідження. Його обирають зі списку, заповнюючи в репозитарії відомості про місце роботи автора, і тоді дослідження однозначно пов'язується з установою (наприклад, з ІЦО НАПН України) незалежно від того, як її назву написано українською чи англійською мовою.



Практична порада: перед першою публікацією даних переконайтеся, що маєте ORCID iD, і дізнайтеся ROR ID своєї установи на сайті ror.org. Тоді DOI набору даних, ORCID iD автора і ROR ID установи будуть пов'язані між собою автоматично.

1.10.3. Підготовка заяви про доступність даних (Data Availability Statement)

Дедалі більше наукових журналів, зокрема ті, що публікують результати педагогічних досліджень, вимагають супроводжувати рукопис публікації заявою про доступність даних (Data Availability Statement) – коротким структурованим текстом, який зазначає, чи були використані в дослідженні оприлюднені дані, де саме (з посиланням на постійний ідентифікатор, розглянутий у підрозділі 1.10.2), на яких умовах доступу (підрозділ 1.1.3) та за наявності яких причин доступ до них обмежено. На відміну від розділу «Подяки» чи вступного посилання на репозитарій у тексті статті, стандартизована заява про доступність даних дає змогу видавничим і бібліометричним системам автоматично встановлювати зв'язок між публікацією та масивом даних, розглянутий у підрозділі 2.2.3, а рецензентам – одразу перевірити доступність матеріалів, на яких ґрунтуються висновки дослідження. Для масивів із обмеженим доступом типове формулювання заяви пояснює причину обмеження (наприклад, наявність персональних даних учасників освітнього процесу, підрозділ 1.9.1) і зазначає процедуру звернення за доступом (підрозділ 1.8.2), а не обмежується констатацією «дані доступні за запитом» без подальших деталей.

1.10.4. Завантаження та верифікація даних у репозитарії

Типовий алгоритм завантаження, спільний для більшості репозитаріїв, розглянутих у підрозділі 1.7.1, охоплює такі кроки: авторизацію через обліковий запис, пов'язаний з ORCID iD; створення нового запису масиву даних; завантаження запакованого об'єкта разом із супровідною документацією; послідовне заповнення обов'язкових і рекомендованих полів метаданих (підрозділ 1.4); вибір ліцензії й режиму доступу, узгодженого з рівнем чутливості даних (підрозділи 1.8.1, 1.9.1); попередній перегляд запису; остаточну публікацію з автоматичним присвоєнням DOI.

Після публікації масиву даних досліднику рекомендується виконати верифікацію його фактичної відповідності принципам FAIR – не лише формально впевнитися в наявності обов'язкових полів, а й перевірити якість реалізації кожного принципу. Для орієнтовної автоматизованої оцінки застосовується інструмент F-UJI (підрозділ 1.4.4). Для глибшої, галузево релевантної оцінки, що враховує специфіку освітніх даних, доцільно застосувати шкалу індикаторів моделі стандартизації FAIR-даних для галузі освітніх наук, розглянуту в підрозділі 1.4.3: на відміну від узагальнених автоматизованих інструментів, вона враховує наявність контрольованих словників, галузевих стандартів і повноту опису педагогічного контексту дослідження.

Верифікація FAIR-відповідності не є одноразовою дією, виконаною відразу після публікації: дослідникові варто періодично перевіряти чинність постійного посилання, актуальність контактної інформації для запитів доступу до обмежених даних та відповідність метаданих останній версії пов'язаної публікації. Лише системний супровід масиву даних протягом усього його життєвого циклу, розглянутого в підрозділі 1.3.1, перетворює формальне дотримання вимог репозитарію на реальну, стійку відповідність принципам FAIR, викладеним у підрозділі 1.1 цього довідника.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ДО РОЗДІЛУ 1



Література до 1.1.

1. Budapest Open Access Initiative. February 14, 2002. URL: https://www.budapestopenaccessinitiative.org/read/
2. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. October 22, 2003. URL: https://openaccess.mpg.de/Berlin-Declaration
3. The FAIR Guiding Principles for scientific data management and stewardship / M. D. Wilkinson et al. *Scientific Data.* 2016. Vol. 3. № 160018. DOI: https://doi.org/10.1038/sdata.2016.18
4. Cost-benefit analysis for FAIR research data - Cost of not having FAIR research data / Directorate-General for Research and Innovation, PwC EU Services. Luxembourg: *Publications Office of the European Union*, 2018. 48 p. DOI: https://doi.org/10.2777/02999
5. The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments / C. Bahim et al. *Data Science Journal.* 2020. Vol. 19. № 41. P. 1–7. DOI: https://doi.org/10.5334/dsj-2020-041
6. Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data. *Luxembourg: Publications Office of the European Union*, 2018. 78 p. DOI: https://doi.org10.2777/1524
7. Іванова С. М., Кільченко А. В. Генеза понятійного концепту з формування та використання FAIR-даних у галузі освітніх наук. *Освіта. Інноватика. Практика.* 2026. Вип.14 (4).С. 28–35. DOI: https://doi.org/10.31110/2616-650X-vol14i4-004
8. Кремень В. Г., Луговий В. І., Регейло І. Ю., Базелюк Н. В., Базелюк О. В. Відкритість, цифровізація й оцінювання в науці: загальне і особливе для соціогуманітарного знання. *Інформаційні технології і засоби навчання*. 2020. Т. 80, № 6. С. 243–266. DOI: https://doi.org/10.33407/itlt.v80i6.4155
9. UNESCO Recommendation on Open Science. *Paris: UNESCO*, 2023. 34 p. DOI: https://doi.org/10.54677/MNMH8546
10. Ярошенко Т., Чуканова С. Принципи FAIR у науці: формування компетенцій для належного управління даними досліджень. *Цифрова платформа: інформаційні технології в соціокультурній сфері*. 2025. Т. 8 (1). С. 223–248. DOI: https://doi.org/10.31866/2617-796X.8.1.2025.335555
11. Hauschke C., Nazarovets S., Altemeier F., Kaliuzhna N. Roadmap to FAIR Research Information in Open Infrastructures. *Journal of Library Metadata*. 2021. Vol. 21. № 1–2. P. 45–61. DOI: https://doi.org/10.1080/19386389.2021.1999156

12. Методичні рекомендації щодо управління науковими даними для закладів вищої освіти та наукових установ у частині визначення механізмів збереження та повторного використання наукових даних. Київ: МОН України, 2024. URL: https://surl.li/xbsswq

13. Олексюк В. П., Спірін О. М., Осадча К. П. Модель стандартизації FAIR-даних для галузі освітніх наук. *Освітня аналітика України.* 2026. Вип. 2. С. 91–103. DOI: https://doi.org/10.32987/2617-8532-2026-2-91-103

14. Dublin Core Metadata Initiative (DCMI). *DCMI Metadata Terms.* URL: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

15. Building school data equity, infrastructure, and capacity through FAIR data standards: Findable, Accessible, Interoperable, and Reusable / A. J. Bowers, Y. Choi. *Educational Researcher.* 2023. Vol. 52. Iss. 7. DOI: https://doi.org/10.3102/0013189X231181103

16. White Paper on implementing the FAIR principles for data in the Social, Behavioural and Educational (SBE) sciences / N. Betancort Cabrera, E. C. Bongartz, N. Dörrenbächer et al. RatSWD Working Paper. No. 274. *German Data Forum,* 2022. URL: https://www.konsortswd.de/wp-content/uploads/RatSWD_WP_274.pdf

17. GO FAIR Initiative. FAIR Principles. URL: https://www.go-fair.org/fair-principles/

18. Вакалюк Т. А., Мінтій І. С. Огляд наявних цифрових сервісів для зберігання й опрацювання FAIR даних в галузі освітніх наук. Наукові записки Бердянського державного педагогічного університету. *Педагогічні науки*, 2026. № 1. С. 357-368. DOI: https://doi.org/10.32782/2412-9208-2026-1-357-368

19. Attribution 4.0 International (CC BY 4.0). *Creative Commons,* 2013. URL: https://creativecommons.org/licenses/by/4.0/

20. Про захист персональних даних: Закон України від 01.06.2010 № 2297-VI. Відомості Верховної Ради України. URL: https://zakon.rada.gov.ua/laws/show/2297-17

21. Voigt P., Von dem Bussche A. The EU General Data Protection Regulation (GDPR): A Practical Guide. *Cham: Springer International Publishing*, 2017. DOI: https://doi.org/10.1007/978-3-319-57959-7

22. RDF 1.1 Concepts and Abstract Syntax / R. Cyganiak, D. Wood, M. Lanthaler (eds.). *W3C Recommendation,* 25 February 2014. URL: https://www.w3.org/TR/rdf11-concepts/

23. OWL 2 Web Ontology Language Document Overview (Second Edition). *W3C Recommendation*, 11 December 2012. URL: https://www.w3.org/TR/owl2-overview/

24. Ontology, Ontologies and the "I" of FAIR / G. Guizzardi. *Data Intelligence.* 2020. Vol. 2. Iss. 1-2. P. 181–191. DOI: https://doi.org/10.1162/dint_a_00040

25. ДСТУ ISO 25964-1:2018. Інформація та документація. Тезауруси та сумісність з іншими словниками. Частина 1. Тезауруси для інформаційного пошуку (ISO 25964-1:2011, IDT). Київ: ДП «УкрНДНЦ», 2019. Чинний від 01.01.2019. https://online.budstandart.com/ua/catalog/doc-page.html?id_doc=80398

26. DDI Alliance. Data Documentation Initiative: metadata standards for the social, behavioral, economic and health sciences. URL: https://ddialliance.org/

27. IEEE Standard for Learning Object Metadata. IEEE Std 1484.12.1-2020. *Institute of Electrical and Electronics Engineers,* 2020. URL: https://standards.ieee.org/ieee/1484.12.1/7699/

28. Experience API (xAPI) Specification. Advanced Distributed Learning (ADL) Initiative. URL: https://github.com/adlnet/xAPI-Spec

29. FAIR Principles: Interpretations and Implementation Considerations / A. Jacobsen, R. de Miranda Azevedo, N. Juty et al. *Data Intelligence.* 2020. Vol. 2. Iss. 1-2. P. 10–29. DOI: https://doi.org/10.1162/dint_r_00024

30. FAIR Competence Framework for Higher Education (Data Stewardship Professional Competence Framework) / Y. Demchenko et al. *FAIRsFAIR Project Deliverable D7.3. Zenodo,* 2021. DOI: https://doi.org/10.5281/zenodo.4562089

Література до 1.2.

1. Осадча К. П., Спірін О. М., Олексюк В. П. Таксономія FAIR-даних для наукових досліджень у галузі освіти в умовах цифрової трансформації. *Вісник Глухівського НПУ ім. О. Довженка,* 2026. Вип. 2 (61). С. 10–22. DOI: https://doi.org/10.31376/2410-0897-2026-2-61-10-22

2. DDI Alliance. Data Documentation Initiative: metadata standards for the social, behavioral, economic and health sciences. URL: https://ddialliance.org/

3. Braun V., Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology.* 2006. Vol. 3. Iss. 2. P. 77-101. DOI: https://doi.org/10.1191/1478088706qp063oa

4. The REFI-QDA Standard. *Rotterdam Exchange Format Initiative*. URL: https://www.qdasoftware.org

5. Taguette – the free and open-source qualitative data analysis tool. URL: https://www.taguette.org

6. Creswell J. W., Plano Clark V. L. Designing and Conducting Mixed Methods Research. 3rd ed. Thousand Oaks, *CA: SAGE Publications*, 2018. 492 p.

7. Про захист персональних даних: Закон України від 01.06.2010 № 2297-VI. Відомості Верховної Ради України. URL: https://zakon.rada.gov.ua/laws/show/2297-17

8. UNESCO Institute for Statistics (UIS). Education Data. URL: https://uis.unesco.org

9. OECD. Education at a Glance. URL: https://www.oecd.org/en/publications/education-at-a-glance_19991487.html

10. Сидоренко В. Методологія і методи наукових досліджень: навчально-методичний посібник. Біла Церква: БІНПО, 2024. 150 с.

Література до 1.3.

1. Higgins S. The DCC Curation Lifecycle Model. *The International Journal of Digital Curation.* 2008. Vol. 3. Iss. 1. P. 134–140. DOI: https://doi.org/10.2218/ijdc.v3i1.48

2. UK Data Service. Roles and Responsibilities. URL: https://ukdataservice.ac.uk/learning-hub/data-producer-support/planning-and-governing-research-data/governance-roles-and-responsibilities/roles-and-responsibilities/

3. Oladipo F., Folorunso S., Ogundepo E., Osigwe O., Akindele A. Curriculum Development for FAIR Data Stewardship. *Data Intelligence.* 2022. Vol. 4. Iss. 4. P. 991–1012. DOI: https://doi.org/10.1162/dint_a_00183

4. FAIR Competence Framework for Higher Education (Data Stewardship Professional Competence Framework) / Y. Demchenko et al. *FAIRsFAIR Project Deliverable D7.3. Zenodo*, 2021. DOI: https://doi.org/10.5281/zenodo.4562089

5. CRediT – Contributor Roles Taxonomy. ANSI/NISO Z39.104-2022. National Information Standards Organization (NISO). URL: https://credit.niso.org/

Література до 1.4.

1. Dublin Core Metadata Initiative (DCMI). DCMI Metadata Terms. URL: https://www.dublincore.org/specifications/dublin-core/dcmi-terms

2. DataCite Metadata Working Group. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.7. *DataCite e.V.,* 2026. DOI: https://doi.org/10.14454/qdd3-ps68

3. FAIR-Aware: Assess Your Knowledge of FAIR (v1.0.1) / V. Akerman, L. Cepinskas, M. Verburg, M. Mustapha. *Zenodo*, 2021. DOI: https://doi.org/10.5281/zenodo.5084861

4. Schema.org. Schema.org Vocabulary. URL: https://schema.org

5. IEEE Standard for Learning Object Metadata. IEEE Std 1484.12.1-2020. *Institute of Electrical and Electronics Engineers*, 2020. URL: https://standards.ieee.org/ieee/1484.12.1/7699/

6. 1EdTech Consortium. About 1EdTech (formerly IMS Global Learning Consortium). URL: https://www.1edtech.org

7. Dublin Core Metadata Initiative (DCMI). *Learning Resource Metadata Initiative (LRMI).* URL: https://www.dublincore.org/resources/lrmi

8. Олексюк В. П., Спірін О. М., Осадча К. П. Модель стандартизації FAIR-даних для галузі освітніх наук. *Освітня аналітика* України. 2026. Вип. 2. С. 91–103. DOI: https://doi.org/10.32987/2617-8532-2026-2-91-103

9. Bhat A., Wani Z. A. The FAIRification process for data stewardship: A comprehensive discourse on the implementation of the FAIR principles for data visibility, interoperability and management. *IFLA Journal.* 2025. Vol. 51. Iss. 2. P. 382–398. DOI: https://doi.org/10.1177/03400352241270692

10. Hellstrand M., Raghothama J., Meijer S. FAIR for whom? A reality check on the state of FAIR research data management in a collaborative research project. *Big Data & Society*. 2025. Vol. 12. Iss. 2. DOI: https://doi.org/10.1177/20539517251349157

Література до 1.5.

1. Data Management Planning: How Requirements and Solutions are Beginning to Converge / S. Jones, R. Pergl, R. Hooft et al. *Data Intelligence.* 2020. Vol. 2. Iss. 1-2. P. 208–219. DOI: https://doi.org/10.1162/dint_a_00043

2. Miksa T., Simms S., Mietchen D., Jones S. Ten principles for machine-actionable data management plans. *PLOS Computational Biology.* 2019. Vol. 15. Iss. 3. Article e1006750. DOI: https://doi.org/10.1371/journal.pcbi.1006750

3. Horizon Europe Data Management Plan Template. Version 1.0, 5 May 2021. European Commission, *Funding & Tenders Portal.* URL: https://ec.europa.eu/info/funding-

tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx

4. Методичні рекомендації щодо управління науковими даними для закладів вищої освіти та наукових установ / Міністерство освіти і науки України. Київ, 2024. URL: https://mon.gov.ua/static-objects/mon/sites/1/nauka/2024/12/31/metod-rekomendatsiyi-shchodo-upravlinnya-naukovymy-danymy-31-12-2024.pdf

5. Practical Guide to the International Alignment of Research Data Management. Extended Edition. *Brussels: Science Europe*, 2021. 52 p. URL: https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf

Література до 1.6.

1. Practical Guide to the International Alignment of Research Data Management. Extended Edition. *Brussels: Science Europe*, 2021. 52 p. URL: https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf

2. ISO 14721:2025. Space Data System Practices – Reference model for an open archival information system (OAIS). *3rd ed. Geneva: International Organization for Standardization*, 2025. URL: https://www.iso.org/standard/87471.html

3. Recommended Formats Statement 2025–2026. *Washington: Library of Congress,* 2025. URL: https://www.loc.gov/preservation/resources/rfs

4. UK Data Service. Format Your Data: Recommended Formats. URL: https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data

Література до 1.7.

1. Вакалюк Т. А., Мінтій І. С. Огляд наявних цифрових сервісів для зберігання й опрацювання FAIR даних в галузі освітніх наук. Наукові записки Бердянського державного педагогічного університету. *Педагогічні науки.* 2026. № 1. С. 357–368. DOI: https://doi.org/10.32782/2412-9208-2026-1-357-368

2. ICPSR – Inter-university Consortium for Political and Social Research. URL: https://www.icpsr.umich.edu

3. TIMSS 2023 International Database / B. Fishbein, M. Taneva, K. Kowolik. Boston College, *TIMSS & PIRLS International Study Center,* 2025. URL: https://timss2023.org/data

4. OECD. PISA Data. URL: https://www.oecd.org/en/about/programmes/pisa/pisa-data.html

5. re3data.org – Registry of Research Data Repositories. URL: https://www.re3data.org

6. FAIRsharing.org – Registry of Standards, Databases and Data Policies. URL: https://fairsharing.org

7. Мінтій І., Вакалюк Т. Аналіз інституційних репозитаріїв та бібліотечних систем для зберігання й опрацювання FAIR-даних у галузі освітніх наук. *Засоби навчальної та науково-дослідної роботи.* 2026. Вип. 66. С. 70–83. DOI: https://doi.org/10.34142/2312-1548.2026.66.04

8. Національний репозитарій академічних текстів (НРАТ). УкрІНТЕІ. URL: https://nrat.ukrintei.ua

9. Електронна бібліотека НАПН України. URL: https://lib.iitta.gov.ua

10. FAIRVault: An Interuniversity Project on Creating a Restricted Access Data Archive / N. Daniels, J. Laureyns, K. Leonard et al. iPRES 2024, *Ghent, Belgium. Zenodo,* 2024. DOI: https://doi.org/10.5281/zenodo.13641958

11. FAIRVault. User Documentation. URL: https://fairvault.github.io/user-documentation/home

12. European Open Science Cloud (EOSC). European Commission, *Research and Innovation.* URL: https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en

13. European Open Science Cloud (EOSC) EU Node. URL: https://open-science-cloud.ec.europa.eu

14. Ukraine. EOSC Association. URL: https://eosc.eu/tripartite-collaboration/ukraine

Література до 1.8.

1. Attribution 4.0 International (CC BY 4.0). *Creative Commons,* 2013. URL: https://creativecommons.org/licenses/by/4.0/

2. Open Data Commons: Legal Tools for Open Data. *Open Knowledge Foundation.* URL: https://opendatacommons.org/

3. About CC Licenses. Creative Commons. URL: https://creativecommons.org/share-your-work/cclicenses/; CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. *Creative Commons.* URL: https://creativecommons.org/publicdomain/zero/1.0

4. Вакалюк Т. А., Мінтій І. С. Огляд наявних цифрових сервісів для зберігання й опрацювання FAIR даних в галузі освітніх наук. *Наукові записки Бердянського державного педагогічного університету. Педагогічні науки*. 2026. № 1. С. 357–368. DOI: https://doi.org/10.32782/2412-9208-2026-1-357-368

Література до 1.9.

1. Про захист персональних даних: Закон України від 01.06.2010 № 2297-VI. Відомості Верховної Ради України. URL: https://zakon.rada.gov.ua/laws/show/2297-17

2. Voigt P., Von dem Bussche A. The EU General Data Protection Regulation (GDPR): A Practical Guide. *Cham: Springer International Publishing*, 2017. DOI: https://doi.org/10.1007/978-3-319-57959-7

3. Prasser F., Eicher J., Spengler H., Bild R., Kuhn K. A. Flexible Data Anonymization Using ARX – Current Status and Challenges Ahead. *Software: Practice and Experience.* 2020. Vol. 50. Iss. 7. P. 1277–1304. DOI: https://doi.org/10.1002/spe.2812

4. Amnesia – Data Anonymization Made Easy. OpenAIRE. URL: https://amnesia.openaire.eu

5. Templ M., Kowarik A., Meindl B. Statistical Disclosure Control for Micro-Data Using the R Package sdcMicro. *Journal of Statistical Software*. 2015. Vol. 67. Iss. 4. P. 1–36. DOI: https://doi.org/10.18637/jss.v067.i04

6. OpenDP: An Open-Source Suite of Differential Privacy Tools. Harvard University, *Institute for Quantitative Social Science; Microsoft.* URL: https://opendp.org

7. Prasser F., Kohlmayer F. Putting Statistical Disclosure Control into Practice: The ARX Data Anonymization Tool. Medical Data Privacy Handbook / A. Gkoulalas-Divanis, G. Loukides (eds.). *Cham: Springer,* 2015. P. 111–148. DOI: https://doi.org/10.1007/978-3-319-23633-9_6

8. OpenDP Library Documentation. OpenDP. URL: https://docs.opendp.org/

9. RO-Crate Metadata Specification 1.3. Research Object Crate community, 2026. URL: https://w3id.org/ro/crate/1.3

10. Kunze J., Littman J., Madden E., Scancella J., Adams C. The BagIt File Packaging Format (V1.0). RFC 8493. *Internet Engineering Task Force*, October 2018. DOI: https://doi.org/10.17487/RFC8493

Література до 1.10.

1. Kunze J., Littman J., Madden E., Scancella J., Adams C. The BagIt File Packaging Format (V1.0). RFC 8493. *Internet Engineering Task Force*, October 2018. DOI: https://doi.org/10.17487/RFC8493

РОЗДІЛ II

ВИКОРИСТАННЯ FAIR-ДАНИХ У ГАЛУЗІ ОСВІТНІХ НАУК

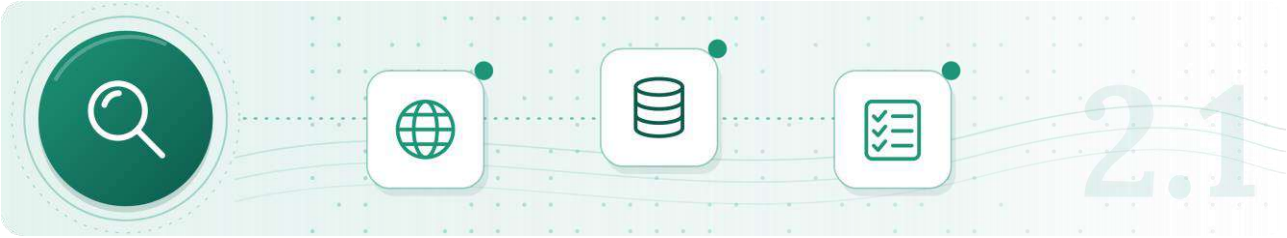


Якщо розділ I довідника розглядав формування FAIR-даних – планування, стандартизацію, зберігання та публікацію масивів даних дослідником-депонентом, то цей розділ звертається до зворотного, але нерозривно пов'язаного з ним процесу: використання вже оприлюднених FAIR-даних іншим дослідником.

Використання FAIR-даних у галузі освітніх наук визначається як сукупність методологічно обґрунтованих і етично відповідальних практик пошуку, оцінювання, доступу, інтерпретації та застосування освітніх дослідницьких даних і метаданих, організованих відповідно до принципів FAIR, з метою верифікації наукових результатів, вторинного аналізу, порівняльних і метааналітичних досліджень, розроблення доказової освітньої політики та вдосконалення педагогічних практик [1].

Процес використання охоплює послідовні етапи: пошук необхідних даних за їх описом, отримання доступу відповідно до встановлених умов, оцінювання якості та надійності даних, поєднання з іншими джерелами, аналітичне опрацювання та коректне цитування, які й визначають структуру цього розділу.

2.1. Пошук та доступ до FAIR-даних



2.1.1. Стратегії пошуку наукових даних (Google Dataset Search, DataCite Search, BASE, OpenAIRE)

Пошук наявних масивів даних є першим етапом їх використання й водночас практичною перевіркою реалізації принципу Findable дослідником, який публікував ці дані (підрозділ 1.1.2). Найбільш масштабне емпіричне дослідження практик пошуку й повторного використання наукових даних, що охопило 1677 респондентів зі 105 країн, засвідчило: дослідники найчастіше розпочинають пошук даних через загальні пошукові системи й лише згодом звертаються до спеціалізованих реєстрів, а визначальними критеріями релевантності знайденого масиву є його тематична відповідність, повнота документування та можливість перевірити метод збору даних [2].

Реєстри й агрегатори, розглянуті в підрозділі 1.7.2, re3data.org і FAIRsharing.org, залишаються основним інструментом цілеспрямованого пошуку профільного репозитарію за предметною галуззю. Доповнюють їх агрегатори прямого пошуку масивів даних: Google Dataset Search індексує масиви даних, розмічені за словником Schema.org (підрозділ 1.4.1), і є найбільш широким за охопленням джерелом завдяки автоматичному вебскануванню; DataCite Search забезпечує пошук за метаданими, зареєстрованими членами консорціуму DataCite, тобто безпосередньо за масивами даних, яким присвоєно DOI (підрозділ 1.10.2); BASE (Bielefeld Academic Search Engine), що адмініструється Бібліотекою Білефельдського університету, агрегує метадані з понад 11 тисяч джерел через протокол OAI-PMH (підрозділ 1.7.3), охоплюючи як дослідницькі дані, так і публікації; OpenAIRE Explore є офіційним пошуковим порталом дослідницьких результатів, профінансованих програмами ЄС, і забезпечує пошук із фільтрацією за зв’язком масиву даних із конкретним проєктом чи публікацією.

Ефективний пошук освітніх наукових даних передбачає комбіноване застосування контрольованої термінології – тезаурусів і класифікацій, розглянутих у підрозділі 1.1.4 (ISCED, EuroVoc), та вільних ключових слів, оскільки метадані частини масивів, особливо розміщених поза

спеціалізованими освітніми сховищами (підрозділ 1.7.2), не завжди послідовно дотримуються контрольованого словника.

2.1.2. Оцінювання якості, прав та приватності знайдених даних

Виявлення масиву даних не є підставою для автоматичної довіри до нього: перш ніж застосувати чужі дані у власному дослідженні, дослідник має виконати процедуру належної перевірки (reuse due diligence), тобто зробити систематичне оцінювання придатності масиву даних до повторного використання за кількома незалежними критеріями. Емпіричні дослідження довіри користувачів до чужих даних засвідчують, що судження про надійність масиву формується на основі декількох джерел: репутації автора й установи, повноти супровідної документації, наявності рецензування чи кураторського контролю під час депонування (підрозділ 1.7.1) та узгодженості метаданих із самими файлами даних [3].

Щоб переконатися, що знайдений масив даних можна правомірно й коректно використати у власному дослідженні, доцільно перед початком роботи з ним відповісти на запитання контрольного переліку, наведеного в табл. 2.1.

Таблиця 2.1.

Контрольний перелік для перевірки масиву даних перед повторним використанням

Критерій	Питання для перевірки
Ліцензія й умови використання	Яку ліцензію присвоєно масиву (підрозділ 1.8.1)? Чи дозволяє вона запланований спосіб використання, зокрема публікацію похідних результатів?
Згода учасників	Чи охоплює інформована згода, зазначена в метаданих або супровідній документації, запланований спосіб повторного використання (підрозділ 1.9.1)?
Ризик повторної ідентифікації	Якщо дані стосуються учасників освітнього процесу, чи застосовано анонімізацію або псевдонімізацію (підрозділ 1.9.2)? Чи не дасть змоги встановити особу поєднання масиву з іншими джерелами? (Псевдонімізовані дані залишаються персональними.)
Повнота документування	Чи супроводжується масив кодовою книгою, файлом README або протоколом аналізу (підрозділи 1.2.1–1.2.2), достатніми для правильного розуміння змінних?
Походження й метод збору	Чи описано методику, вибірку та інструментарій дослідження, а також зміни, яких зазнали дані (підрозділи 1.1.5, 2.2.3)? Чи відповідають вони меті запланованого повторного використання?

Критерій	Питання для перевірки
Технічна сумісність	У якому форматі подано дані (підрозділ 1.6)? Чи потрібне перетворення в інший формат і чи не буде при цьому втрачено частину інформації?
Цілісність і версія	Яку версію масиву використано і чи є вона останньою? Чи збігаються контрольні суми файлів, якщо репозитарій їх надає (підрозділ 1.6.1)?
Цитування	Чи вказано рекомендований спосіб цитування та DOI масиву (за наявності – DOI конкретної версії)?

Складено за [3] із доповненням авторів.

Практичні настанови провідних університетських служб управління дослідницькими даними узагальнюють ці критерії в три послідовні принципи належної перевірки: допитливість – вивчити масив даних до повного розуміння його придатності для запланованого дослідження; критичність – застосовувати наведені вище критерії оцінювання, а не приймати метадані на віру; дотримання – неухильно виконувати умови ліцензії, згоди й угод про використання даних, навіть якщо технічний доступ до масиву не обмежений [4].

2.1.3. Правові та етичні аспекти доступу

З погляду дослідника-користувача, а не депонента, правовий і етичний вимір доступу до FAIR-даних зумовлений тими самими нормативними рамками, що розглянуті в підрозділі 1.9, а саме: Законом України «Про захист персональних даних» та GDPR, однак застосовується до нової мети опрацювання. Принциповим є те, що згода, надана учасниками освітнього процесу на етапі первинного дослідження, не поширюється автоматично на будь-яке подальше використання даних: дослідник, який має намір застосувати чужий масив для нової дослідницької мети, зобов'язаний перевірити, чи охоплює первинна згода цю мету (підрозділ 1.9.1), а за потреби звернутися за додатковим дозволом до розпорядника даних або комітету з етики установи-депонента (підрозділ 1.3.2).

Для масивів даних із обмеженим режимом доступу, розглянутим у підрозділі 1.1.3, процедура доступу може передбачати підписання угоди про використання даних (Data Use Agreement), що визначає дозволені аналітичні операції, заборону спроб повторної ідентифікації учасників і зобов'язання знищити отримані дані після завершення дослідження. За архітектурою, подібною до розглянутого в підрозділі 1.7.5 сервісу FAIRVault, рішення про

надання доступу до чутливих освітніх даних ухвалює комітет з доступу до даних установи-розпорядника, а не автоматизована система, що узгоджується з диференціацією автоматизованих і ручних процедур автентифікації (підрозділ 1.8.2).

2.2. Технічна і семантична інтеграція FAIR-даних у дослідженні



2.2.1. Узгодження форматів і термінології даних з різних джерел

Знайдений і перевірений набір даних (підрозділ 2.1) рідко вдається одразу використати у власному дослідженні: він може бути збережений в іншому форматі, мати інші назви змінних або іншу термінологію. Тому перед аналізом чужі дані потрібно узгодити з власними – привести до спільного формату і спільних позначень.

Для табличних кількісних даних, найпоширенішого типу освітніх даних (підрозділ 1.2.1), головним засобом узгодження є опис кожної змінної: її назва, формулювання запитання, тип і шкала значень, одиниця вимірювання, допустимі значення та позначення пропусків. Такий опис міститься в кодовій книзі. Консорціум World Wide Web (W3C) розробив також рекомендацію CSVW (CSV on the Web), яка дає змогу подати опис стовпців таблиці в машиночитаному вигляді окремим файлом поряд із самою таблицею [1]; це спрощує автоматичну перевірку узгодженості даних із кількох джерел, однак для більшості педагогічних досліджень достатньо якісної кодової книги.

Поєднуючи дані з кількох джерел, дослідник зіставляє змінні: з'ясовує, чи однаково в різних наборах закодовано, наприклад, рівень освіти, стать, шкалу оцінювання чи тип закладу, і за потреби перекодовує їх до спільного класифікатора – ISCED для рівнів освіти, EuroVoc чи освітніх тезаурусів для тематики (підрозділ 1.1.4). Наприклад, якщо в одному наборі вказано «9 клас», а в іншому – «базова середня освіта», обидва значення зводять до відповідного рівня ISCED. Усі такі рішення фіксують у документації – найзручніше у вигляді таблиці відповідності змінних, яку публікують разом з об'єднаним набором даних. Автоматизоване зіставлення даних на основі онтологій (RDF, DCAT)

можливе лише тоді, коли первинні набори вже описано контрольованими словниками, і є завданням інформаційних фахівців.

2.2.2. Інтеграція зовнішніх даних у власне дослідження

Дослідники, що застосовують чужі масиви даних, вдаються до принципово різних способів їх залучення до власного дослідження. У класичній типології повторного використання наукових даних виокремлено *чотири типи*: 1) верифікація – відтворення результатів первинного дослідження на тих самих даних для перевірки коректності висновків; 2) реплікація – застосування тієї самої методики до нового, самостійно зібраного масиву даних; 3) розширення – нові аналізи на основі наявного масиву, які первинні автори не проводили; 4) рекомбінація – поєднання кількох незалежних масивів даних у єдиний набір для аналізу, що жоден з окремих масивів не уможливлював [2]. Кожен із цих типів висуває різні вимоги до технічної інтеграції: верифікація й реплікація потребують точної відповідності формату й структури первинного масиву, тоді як рекомбінація, найскладніший із чотирьох типів, вимагає розв'язання всіх завдань семантичного узгодження (підрозділ 2.2.1).

Практична інтеграція масивів різного походження, наприклад, поєднання результатів власного опитування з відкритими адміністративними даними закладу освіти (підрозділ 1.2.4) чи з мікроданими міжнародного порівняльного дослідження (підрозділ 1.7.2), вимагає визначення спільного ключа зв'язування записів: часової позначки, коду закладу освіти чи (за наявності належної анонімізації, розглянутої в підрозділі 1.9.2) псевдонімізованого ідентифікатора учасника. Якість такої інтеграції безпосередньо залежить від повноти документування первинних масивів – кодової книги чи протоколу аналізу, розглянутих у підрозділах 1.2.1-1.2.2, без яких зіставлення змінних різних джерел перетворюється на здогад, а не на відтворювану процедуру.

Дослідження практик повторного використання наукових даних засвідчують, що вирішальну роль у рішенні використати чужі дані відіграє довіра до них. Користувачі даних формують її на основі кількох чинників: репутації авторів і репозитарію, повноти документації, прозорості методики збирання даних і можливості зв'язатися з автором для уточнень [3]. Для освітян, які плануватимуть інтегрувати зовнішні дані у власне дослідження, звідси випливає практична порада: перш ніж поєднувати масиви, варто оцінити кожен із них за цими критеріями (підрозділ 2.1.2) і за потреби звернутися до авторів із запитаннями щодо незрозумілих змінних чи процедур.

2.2.3. Цитування, походження даних і відтворюваність дослідження

Коректне цитування масиву даних, використаного повністю, частково чи як методичне підґрунтя власного дослідження, є формою визнання інтелектуального внеску його авторів і водночас необхідною умовою відтворюваності наукових результатів. Стандартна структура цитати масиву даних, аналогічна цитуванню публікації, включає авторів, рік, назву, видавця (репозитарій), версію та постійний ідентифікатор (DOI), розглянутий у підрозділі 1.10.2; саме присутність активного DOI в переліку використаних джерел дає змогу науковим системам, таким як Scopus, Web of Science, DataCite, автоматично встановлювати зв'язок між масивом даних і публікаціями, які його використали. Але попри існування стандарту цитування даних понад десятиліття, значна частка публікацій продовжує згадувати використані масиви лише в тексті, без формального цитування з ідентифікатором, що ускладнює облік використання даних і применшує визнання роботи їх авторів [4].

Масштабний аудит відтворюваності наукових публікацій – огляд 351 статті провідного видання з географічних інформаційних систем за 2020-2024 рр., опублікованих після запровадження вимоги відкривати дані та програмний код, засвідчив, що частка статей із доступними матеріалами істотно зросла, однак значна частина досліджень залишається практично невідтворюваною через неповний опис того, як саме опрацьовували дані, а не через відсутність самих даних [5].

Відповідно до Спільної декларації принципів цитування даних, підготовленої міжнародною ініціативою FORCE11, дані слід розглядати як повноцінний результат наукової діяльності, а їх цитування має забезпечувати визнання авторів, однозначну ідентифікацію використаних даних і постійний доступ до них. Декларація містить вісім принципів: важливість, визнання й атрибуція, доказовість, унікальна ідентифікація, доступ, сталість, точність і перевірюваність, інтероперабельність і гнучкість [6]. Принцип точності, зокрема, означає, що в посиланні слід зазначати конкретну версію чи частину набору даних, яку використано в дослідженні.

Для галузі освітніх наук звідси можна зробити важливий висновок: відкритий набір даних є необхідною, але не достатньою умовою відтворюваності дослідження. Вирішальне значення має документування походження даних (підрозділ 1.1.5), послідовний опис усіх кроків від первинних даних до результатів аналізу: які записи вилучено й чому, як перекодовано змінні, які методи й програми застосовано.

2.3. Аналіз, вторинне використання та метадослідження на основі FAIR-даних



Технічно й семантично інтегровані дані, розглянуті в підрозділі 2.2, стають підставою для чотирьох напрямів наукового застосування, визначених у концептуальному апараті використання FAIR-даних у галузі освітніх наук [1]: вторинного аналізу, метааналітичних досліджень, міжнародних порівнянь та розроблення доказової освітньої політики. Цей підрозділ послідовно розглядає кожен із них.

2.3.1. Вторинний аналіз освітніх даних, зокрема методами машинного навчання

Вторинний аналіз – це застосування нових методів до масиву даних, зібраного для інших дослідницьких цілей, – дедалі ширше використовується в педагогічних дослідженнях, зокрема завдяки методам машинного навчання, здатним виявляти складні закономірності у великих масивах даних. Показовим прикладом є опубліковане у 2026 р. дослідження, у якому методами машинного навчання проаналізовано дані Міжнародного дослідження викладання та навчання TALIS 2024 (відповіді 88991 вчителя з 55 країн і територій) і встановлено зв’язок між використанням засобів штучного інтелекту під час підготовки до уроків і тим, наскільки, за власною оцінкою вчителів, вони досягають цілей уроку [2]. Це дослідження ілюструє головну перевагу вторинного аналізу: дані, зібрані для одного завдання (моніторингу практики викладання), дали змогу відповісти на нове запитання, якого не ставили первинні автори, – саме завдяки тому, що масив TALIS опубліковано з докладною документацією у відкритих форматах.

Вторинний аналіз даних міжнародних порівняльних досліджень (PISA, TIMSS, PIRLS, TALIS) має свої методологічні особливості, які необхідно враховувати навіть досвідченим дослідникам. У цих дослідженнях застосовують складні багатоступеневі вибірки, тому для отримання коректних оцінок слід використовувати вибіркові ваги; результати учнів подаються не одним балом, а набором так званих правдоподібних значень (plausible values),

які потрібно аналізувати разом; стандартні похибки обчислюють спеціальними методами повторних вибірок. Ігнорування цих особливостей, як показують методологічні огляди, призводить до хибних висновків і є однією з найпоширеніших помилок вторинного аналізу таких даних [3].

Щоб полегшити роботу з такими даними, організатори досліджень надають безкоштовні інструменти аналізу. Зокрема, Міжнародна асоціація з оцінювання освітніх досягнень (IEA) пропонує програму IDB Analyzer, яка автоматично враховує вибіркові ваги й правдоподібні значення під час аналізу даних IEA та інших масштабних досліджень, а також портал ILSA Gateway з доступом до даних і документації міжнародних досліджень [4]. Для освітян, які лише починають працювати з цими даними, доцільно розпочинати з відтворення результатів, наведених в офіційних звітах: збіг отриманих значень з опублікованими свідчить про правильність застосованої методики.

Вторинний аналіз відкритих даних має ще одну перевагу для освітніх досліджень: він не потребує повторного залучення учасників, що зменшує навантаження на заклади освіти, вчителів і здобувачів освіти, яких нерідко запрошують до численних опитувань. Водночас дослідник, який працює з чужими даними, має чесно зазначити в публікації, що дані зібрано іншими авторами, процитувати набір даних (підрозділ 2.2.3) і описати обмеження, зумовлені тим, що первинне дослідження не планувалося для відповіді на нове запитання.

Придатність масиву освітніх даних до вторинного аналізу визначається якістю його підготовки на етапі формування (розділ 1 довідника): узгодженою табличною структурою (підрозділ 1.6.2), мінімальною кількістю пропусків і достатньою документацією для коректного розуміння змінних (підрозділи 1.2.1–1.2.2). Водночас, застосовуючи методи машинного навчання до даних про учасників освітнього процесу, навіть анонімізованих (підрозділ 1.9.2), слід бути обачними: за певних умов результати таких моделей можуть опосередковано розкрити інформацію про окремих осіб, тому для чутливих даних рішення про такий аналіз варто погоджувати з фахівцем із захисту персональних даних (підрозділ 1.3.2).

2.3.2. Методи кількісного й якісного аналізу; інструменти освітньої аналітики

Аналітичне опрацювання FAIR-даних застосовує ті самі методи кількісного й якісного аналізу, що й у первинному дослідженні (підрозділи 1.2.1–1.2.2), однак дослідник, який працює з чужими даними, обмежений методологічними рішеннями, ухваленими на етапі первинного збору, зокрема

обраною шкалою вимірювання чи структурою запитань анкети, які вже неможливо змінити. Це висуває додаткову вимогу до попереднього ознайомлення з кодовою книгою чи протоколом аналізу (підрозділи 1.2.1–1.2.2) ще до початку власне аналітичного етапу. Окрему категорію застосунків становить освітня аналітика (learning analytics) – аналіз журналів активності здобувачів освіти в цифрових навчальних середовищах (підрозділ 1.2.1), що дедалі частіше поєднує описову статистику з предиктивним моделюванням для завчасного виявлення здобувачів у зоні ризику неуспішності чи відрахування. Інструментарій освітньої аналітики розвивається у напрямі стандартизованих специфікацій даних про навчальну активність, зокрема xAPI, розглянутого в підрозділі 1.1.4, що дає змогу застосовувати той самий аналітичний конвеєр до даних, отриманих із різних платформ управління навчанням.

Загальновживане визначення навчальної аналітики, запропоноване на першій міжнародній конференції з навчальної аналітики та знань (LAK, 2011), розглядає її як вимірювання, збирання, аналіз і подання даних про здобувачів освіти та контексти їхнього навчання з метою розуміння й оптимізації навчання та середовищ, у яких воно відбувається [5]. Поряд із навчальною аналітикою розвивається суміжний напрям – інтелектуальний аналіз освітніх даних (educational data mining), який більше зосереджений на автоматичному виявленні закономірностей у даних. Огляди досліджень в обох напрямах засвідчують, що найпоширенішими завданнями є прогнозування успішності, виявлення здобувачів, яким загрожує відрахування, аналіз навчальної поведінки та надання зворотного зв’язку викладачам і здобувачам [6].

Для більшості завдань вторинного аналізу освітніх даних не потрібне складне програмне забезпечення. Описову статистику, порівняння груп, кореляційний і регресійний аналіз, перевірку надійності шкал можна виконати в безкоштовних програмах із графічним інтерфейсом, зокрема jamovi, що поєднує звичний вигляд електронної таблиці з повним набором статистичних методів і зберігає дані, аналіз і результати в одному файлі, що полегшує відтворення дослідження [7]. Для якісного аналізу текстів: відкритих відповідей анкет, транскриптів інтерв’ю, можна використовувати безкоштовний інструмент Taguette (підрозділ 1.2.2). Складніші методи, зокрема машинного навчання, як правило, потребують мов програмування R чи Python і співпраці з фахівцями з аналізу даних.

Обираючи метод аналізу чужих даних, дослідник має зважати на те, як ці дані було зібрано. Наприклад, якщо вибірка не є репрезентативною, результати не можна поширювати на всю сукупність вчителів чи здобувачів освіти; якщо

шкала вимірювання порядкова, некоректно обчислювати для неї середнє значення без застереження. Ці обмеження мають бути описані в кодовій книзі чи README первинного набору даних (підрозділи 1.2.1, 1.10.1), а дослідник, який здійснює вторинний аналіз, зобов'язаний зазначити їх у власній публікації.

Важливою вимогою до аналітичного опрацювання FAIR-даних є прозорість: у публікації чи разом із нею варто оприлюднити скрипт або детальний опис усіх кроків аналізу, щоб інші дослідники могли їх відтворити (підрозділ 2.2.3). Програми, які зберігають історію аналізу (jamovi, R, синтаксис SPSS), полегшують виконання цієї вимоги.

Отже, вибір інструментів аналізу залежить насамперед від дослідницького запитання та типу даних, а не від популярності програми; перевагу варто надавати безкоштовним інструментам із відкритим кодом, оскільки вони дають змогу будь-якому досліднику відтворити аналіз без придбання ліцензій.

2.3.3. Метааналіз, систематичні огляди та міжнародні порівняльні дослідження

Метааналіз означає статистичне узагальнення результатів кількох незалежних досліджень однієї й тієї самої дослідницької проблеми і є формою повторного використання не первинних даних, а розмірів ефекту (effect sizes), обчислених у первинних дослідженнях; водночас якість метааналізу прямо залежить від повноти звітування первинних досліджень, включно з тими характеристиками вибірки й методики, що уможливлюють коректне обчислення розміру ефекту. Масштабний огляд методологічної якості 103 метааналізів шкільних академічних втручань засвідчив: попри вдосконалення статистичних моделей метааналізу за останнє десятиліття, автори систематично недостатньо звітують про характеристики вибірки й контексту первинних досліджень, що обмежує здатність практиків визначити застосовність узагальнених результатів до власного контексту [8]. Це прямо узгоджується з вимогою повноти документування (R1), розглянутою в підрозділі 1.1.5 як умова придатності даних до повторного використання: неповне звітування є однаковою мірою перешкодою і для метааналізу результатів, і для вторинного аналізу самих даних.

Систематичні огляди й метааналізи в освіті ґрунтуються на прозорих і відтворюваних процедурах пошуку та добору досліджень. Міжнародним стандартом звітування про систематичні огляди є настанова PRISMA 2020, яка містить перелік із 27 пунктів і схему, що відображає кількість знайдених,

перевірених, вилучених і включених до огляду публікацій [9]. У контексті FAIR-даних важливо, що сам систематичний огляд також створює дані: перелік знайдених публікацій, таблиці закодованих характеристик досліджень, обчислені розміри ефекту. Оприлюднення цих матеріалів у репозитарії разом із протоколом огляду дає змогу іншим дослідникам перевірити висновки, оновити огляд у майбутньому або використати закодовані дані для нових запитань.

Метааналітичні узагальнення відіграють важливу роль у підготовці практичних рекомендацій для освітян. Наприклад, вже згаданий огляд методологічної якості метааналізів шкільних втручань наголошує, що для практиків важливо знати не лише середній ефект, а й для яких учнів, у яких умовах і за якої реалізації втручання він досягається [8]. Тому, плануючи власне дослідження, педагог-науковець може зробити внесок у майбутні метааналізи вже тим, що повно опише вибірку, умови проведення та інструменти і оприлюднить дані відповідно до принципів FAIR.

Для міжнародних порівнянь у галузі освіти важливе значення має також порівнянність показників. Організатори масштабних досліджень спеціально перевіряють, чи однаково розуміють запитання й завдання учасники з різних країн і мовних груп, а результати такої перевірки публікують у технічній документації. Дослідник, який здійснює власне порівняння, має ознайомитися з цією документацією, щоб не робити висновків на основі показників, які в різних країнах вимірюють різне.

Міжнародні порівняльні дослідження – окрема форма використання FAIR-даних, що спирається на масиви даних, спеціально зібрані за уніфікованою методикою в кількох країнах, насамперед TIMSS, PIRLS і PISA (підрозділ 1.7.2). Відкритий доступ до мікроданих цих досліджень, супроводжуваних докладною документацією вибірки й інструментарію, дає змогу дослідникам галузі освітніх наук проводити власні порівняльні аналізи, не обмежені офіційними звітами організаторів дослідження, наприклад, зіставляти зв’язок конкретних освітніх практик із результатами навчання в підгрупах країн, не розглянутих у первинній публікації.

2.3.4. Розроблення доказової освітньої політики на основі FAIR-даних

Розроблення доказової освітньої політики (evidence-based policy) є кінцевою метою значної частини напрямів використання FAIR-даних, розглянутих у цьому підрозділі: верифіковані результати вторинного аналізу, узагальнення метааналізів і висновки міжнародних порівнянь мають цінність для розробників освітньої політики лише за умови, що вони доходять до них у

придатній для використання формі. Систематичний огляд чинників, які визначають застосування наукових результатів в освітній політиці, здійснений Організацією економічного співробітництва та розвитку на основі аналізу дослідницьких публікацій і опитування 288 організацій-посередників знань у 34 країнах, засвідчив: наявність доступних, відкритих масивів даних і результатів дослідження є необхідною, але недостатньою умовою їх застосування в політиці, а визначальну роль відіграє діяльність організацій-посередників, які адаптують наукові результати до формату, зрозумілого й застосовного для органів управління освітою [10].

Міжнародний досвід засвідчує ефективність посередницьких ресурсів, які узагальнюють результати досліджень у зрозумілій для освітян і управлінців формі. Прикладом є інструментарій «Teaching and Learning Toolkit» британського Фонду освітніх досягнень (Education Endowment Foundation), створений у співпраці з Даремським університетом: для кожного педагогічного підходу (зворотний зв'язок, домашні завдання, навчання однолітками тощо) у ньому наведено узагальнену оцінку впливу на навчальні результати, вартості впровадження та надійності доказової бази [11]. Подібні ресурси можливі лише за наявності великої кількості досліджень з відкритими й порівнянними даними, тобто саме тієї інфраструктури, розбудові якої сприяють принципи FAIR.

Прихильники доказового реформування освіти наголошують, що цикл «розроблення – оцінювання – поширення» успішних практик, який давно діє в медицині, може бути перенесений і в освіту за умови систематичного накопичення результатів досліджень та їх доступності для практиків [12]. Для наукових установ НАПН України це відкриває перспективу: публікуючи дані педагогічних експериментів у відкритому доступі з належною документацією, установи не лише виконують вимоги відкритої науки, а й створюють підґрунтя для майбутніх узагальнень, на які зможуть спиратися органи управління освітою.

Водночас використання даних для ухвалення управлінських рішень потребує обережності. Показники, отримані в одному контексті, не завжди можна безпосередньо переносити в інший; агреговані дані можуть приховувати суттєві відмінності між регіонами чи типами закладів освіти; а оприлюднення рейтингів окремих закладів без урахування контексту може мати небажані наслідки. Тому в аналітичних матеріалах для органів управління освітою варто поряд із висновками чітко зазначати обмеження даних і методів, на яких вони

ґрунтуються, та посилання на відкриті набори даних, що дає змогу незалежно перевірити отримані результати.

Для наукових установ і закладів освіти це означає, що публікація FAIR-даних і результатів їх аналізу відповідно до принципів, розглянутих у розділі I довідника, є необхідним, але не самодостатнім кроком: реалізація потенціалу доказової освітньої політики вимагає також цілеспрямованої діяльності з поширення результатів у форматі, доступному органам управління освітою, – коротких аналітичних записок, інтерактивних панелей показників чи безпосередньої участі дослідників у консультативних процесах підготовки нормативних документів.

2.4. Приклади застосування FAIR-даних у педагогічних дослідженнях



Загальні напрями використання FAIR-даних, розглянуті в підрозділі 2.3, конкретизуються по-різному залежно від типу освітнього дослідження. Цей підрозділ розглядає чотири показові приклади: перші три ілюструють специфічні виклики формування й використання FAIR-даних у психолого-педагогічних опитуваннях, навчальній аналітиці та новітній практиці застосування синтетичних даних, а четвертий демонструє покрокову процедуру створення FAIR-пакета.

2.4.1. Опитування та психолого-педагогічне оцінювання

Дані психолого-педагогічних опитувань і стандартизованих психометричних інструментів: шкал мотивації, тривожності, ставлень, оцінювання дидактичних компетентностей, становлять значну частку кількісних освітніх даних, розглянутих у підрозділі 1.2.1, і водночас є категорією, для якої проблема FAIR-відповідності є емпірично добре задокументованою. Спеціальний тематичний випуск журналу Journal of Open Psychology Data, присвячений саме психологічним даним в освітній галузі, наголошує, що доступ до FAIR-сумісних даних у цій сфері залишається істотним викликом, і представляє приклади масивів, підготовлених відповідно до принципів FAIR як зразок для наслідування [1]. Найбільший на сьогодні

емпіричний аналіз стану FAIR-відповідності в психологічній науці – це оцінювання 11384 масивів психологічних даних за 2013–2024 рр. за допомогою інструменту F-UJI, розглянутого в підрозділі 1.4.4, засвідчив цікаву практичну закономірність: масиви даних, як правило, добре відповідають складникам Findable та Interoperable, однак істотно гірше складникам Accessible та Reusable, причому рівень FAIR-відповідності визначається переважно вибором репозитарію, а не роком публікації, підгалуззю психології чи редакційною політикою журналу [2]. Це підтверджує актуальний для освітніх наук висновок: технічна відшукуваність масиву даних не є еквівалентом його практичної придатності до повторного використання, а розрив між цими двома вимірами усувається насамперед вибором належної інфраструктури зберігання (підрозділ 1.7), а не самим фактом публікації.

Психолого-педагогічні методики мають ще одну особливість, яку слід ураховувати під час публікації даних: багато стандартизованих опитувальників і тестів захищено авторським правом. У такому разі дослідник може оприлюднити відповіді учасників і обчислені показники, але не повний текст методики. У кодовій книзі доцільно навести назву інструменту, бібліографічне посилання на нього, номери й короткий зміст пунктів (за дозволом правовласника – і їх формулювання), спосіб підрахунку шкальних балів і показники надійності на вибірці дослідження. Для авторських анкет, розроблених у межах дослідження, варто оприлюднити повний текст інструменту разом із даними – це дає змогу іншим дослідникам застосувати його повторно й порівняти результати.

Як засвідчують спеціалізовані видання з відкритих психологічних даних, якісно задокументований набір даних опитування в галузі освіти містить: опис вибірки та процедури збирання даних, повний перелік змінних з їх значеннями, інформацію про обробку пропущених значень і вилучення записів, посилання на пов’язані публікації та чіткі умови використання [1]. Публікацію такого набору даних в окремому журналі даних (data journal) чи у вигляді статті-опису даних можна розглядати як окремий науковий результат, що підвищує видимість роботи дослідника та полегшує цитування самих даних.

Під час збирання даних за допомогою онлайн-форм (Google Forms, Microsoft Forms, LimeSurvey) варто пам’ятати, що за замовчуванням ці сервіси можуть фіксувати адреси електронної пошти чи інші відомості про респондентів. Якщо опитування передбачає анонімність, ці функції слід вимкнути ще до початку збирання даних, а вивантажений файл перевірити на

наявність ідентифікаційних відомостей перед подальшим опрацюванням і публікацією (підрозділ 1.9.2).

Таким чином, для психолого-педагогічних опитувань принципи FAIR реалізуються через поєднання трьох складників: якісної документації інструменту й шкал, ретельного знеособлення відповідей і вибору репозитарію, який забезпечує повноцінний опис і стабільний доступ до даних.

2.4.2. Навчальна аналітика і захист персональних даних

Навчальна аналітика (learning analytics) – це систематичний аналіз журналів активності здобувачів освіти в цифрових навчальних середовищах (підрозділ 1.2.1). Дані навчальної аналітики становлять окремий клас освітніх даних, для якого напруга між дослідницькою цінністю деталізованих поведінкових даних і вимогами захисту персональних даних (підрозділ 1.9) є особливо гострою: журнали активності за визначенням фіксують дії конкретного здобувача освіти з високою часовою деталізацією, що створює значно вищий ризик повторної ідентифікації, ніж агреговані результати тестування. Систематичний огляд досліджень із питань приватності й захисту даних у навчальній аналітиці констатує, що більшість наявних технологічних рішень зосереджені на технічних аспектах анонімізації, тоді як організаційні й процедурні вимоги – прозоре інформування здобувачів про збір даних, реалізація їхнього права на доступ до власних даних і на заперечення проти автоматизованого прийняття рішень, передбачені статтями 13–15 та 22 GDPR, залишаються систематично недостатньо реалізованими у практиці закладів освіти [3]. Емпіричні дослідження готовності здобувачів освіти ділитися даними навчальної активності також засвідчують, що ця готовність істотно залежить від контексту використання даних і можливості обговорення умов такого використання, а не лише від формальної наявності згоди, що узгоджується з розглянутим у підрозділі 1.9.1 принципом визначеності мети опрацювання персональних даних.

Для закладів освіти, які планують використовувати навчальну аналітику, розроблено практичні інструменти, що допомагають урахувати етичні й правові вимоги. Контрольний перелік DELICATE пропонує вісім запитань, на які варто відповісти перед запровадженням аналітики: визначити мету й переваги її використання, пояснити здобувачам освіти, які дані збираються і для чого, обґрунтувати законність опрацювання, залучити всіх зацікавлених сторін, отримати згоду, знеособити дані, забезпечити технічний захист і врегулювати умови роботи із зовнішніми партнерами [4]. Інший інструмент – рамка SHEILA,

розроблена в межах європейського проєкту на основі досвіду ЗВО, допомагає установі сформувати політику навчальної аналітики, визначивши цілі, відповідальних осіб, правила доступу до даних та порядок оцінювання результатів [5].

Для педагогів, які використовують дані електронних курсів у власних дослідженнях, ці вимоги можна звести до кількох практичних правил. Здобувачі освіти мають заздалегідь знати, які дані про їхню навчальну активність фіксуються і як вони будуть використані. Для дослідження варто вивантажувати лише ті дані, які справді потрібні для відповіді на дослідницьке запитання, і одразу замінювати імена та облікові записи кодами. Результати аналізу не повинні впливати на оцінювання конкретних здобувачів, якщо це не передбачено умовами навчання. А для публікації набору даних журнали активності, як правило, слід агрегувати (наприклад, до кількості переглядів чи середнього часу роботи з ресурсом на тиждень) або надавати лише за обмеженим доступом (підрозділ 1.8.2).

Отже, навчальна аналітика може бути цінним джерелом FAIR-даних для педагогічних досліджень лише за умови, що прозорість для здобувачів освіти, мінімізація даних і контроль доступу закладено в неї від початку, а не додано після того, як дані вже зібрано [6].

2.4.3. Синтетичні дані як інструмент етичного повторного використання

Синтетичні дані – це штучно згенеровані набори, які відтворюють статистичні властивості реального масиву (розподіли значень, зв'язки між змінними), але не містять записів про реальних учасників дослідження. Вони є новим способом розв'язати суперечність між потребою у відкритих, деталізованих освітніх даних і вимогами захисту приватності учасників освітнього процесу (підрозділ 1.9). Для створення синтетичних даних використовують генеративні моделі штучного інтелекту, які нерідко поєднують із диференційною приватністю (підрозділ 1.9.2), щоб обмежити ризик розкриття інформації про конкретного учасника первинного масиву [6]. Прикладом є оприлюднений у 2026 р. синтетичний масив SynEdu-HEDL, що відтворює 20000 записів студентів ЗВО за 85 ознаками – демографічними характеристиками, особливостями навчальної взаємодії в часі та показниками успішності [7].

Водночас дослідники застерігають від сприйняття синтетичних даних як універсального розв'язання проблеми приватності. Рідкісні чи нетипові випадки – наприклад, здобувачі освіти з незвичними навчальними стратегіями –

можуть залишатися впізнаваними навіть у синтетичному масиві, а поєднання синтетичних даних із диференційною приватністю ще потребує перевірки саме в освітньому контексті: висновки, отримані в охороні здоров'я чи фінансах, не можна безпосередньо переносити на освітні дані [8]. Тому установам, які розглядають публікацію синтетичних версій чутливих освітніх масивів як доповнення до обмеженого доступу до первинних даних (підрозділ 1.7.5), варто для кожного типу даних окремо зважувати баланс між захистом приватності та аналітичною цінністю даних (підрозділ 1.9.2).

Для дослідника-освітянина синтетичні дані можуть бути корисними в кількох практичних ситуаціях: для навчання аспірантів і студентів методів аналізу без доступу до реальних персональних даних; для попередньої перевірки скриптів аналізу перед поданням запиту на доступ до справжніх даних з обмеженим доступом; для демонстрації структури набору даних у публікації чи репозитарії, коли первинні дані не можуть бути оприлюднені. В останньому випадку синтетичний набір даних публікують з чітким позначенням, що він не є реальним, а висновки, отримані на ньому, потребують перевірки на первинних даних.

Створення якісних синтетичних даних є технічно складним завданням, яке, як правило, виконують фахівці з аналізу даних. Тому освітянам, які розглядають можливість використання синтетичних даних, доцільно співпрацювати з такими фахівцями, а в описі синтетичного набору даних обов'язково зазначати, на основі якого первинного масиву й якими методами його створено, наскільки добре він відтворює властивості первинних даних і які заходи вжито для захисту приватності учасників (підрозділ 1.9.2).

2.4.4. Покроковий приклад створення FAIR-пакета

Щоб продемонструвати, як положення розділу I поєднуються в цілісну процедуру, розглянемо умовний приклад: дослідник провів онлайн-опитування 120 викладачів закладів вищої освіти щодо використання цифрових інструментів у навчанні й планує оприлюднити відповіді як набір FAIR-даних. Нижче подано послідовність кроків із посиланнями на відповідні підрозділи довідника.

1

Крок 1. План управління даними. До початку опитування дослідник заповнює шаблон плану управління даними (підрозділ 1.5.2): зазначає, які дані буде зібрано (відповіді на закриті та кілька відкритих запитань), у яких форматах їх зберігатимуть, як забезпечуватиметься знеособлення, які репозитарій і ліцензію буде обрано. У формі інформованої згоди учасників повідомляють, що знеособлені відповіді буде оприлюднено для повторного використання.

2

Крок 2. Збирання та знеособлення даних. У налаштуваннях онлайн-форми вимикають збирання адрес електронної пошти (підрозділ 2.4.1). Після вивантаження з файлу вилучають позначки часу та назви закладів освіти, вік і стаж групують в інтервали, а у відповідях на відкриті запитання видаляють імена й назви, за якими можна ідентифікувати респондента; ризик повторної ідентифікації перевіряють за допомогою інструментів, описаних у підрозділі 1.9.2.

3

Крок 3. Підготовка файлів даних і документації. Відповіді зберігають у файлі CSV (кодування UTF-8, перший рядок – короткі назви змінних латиницею). Кодова книга містить для кожної змінної її назву, повне формулювання запитання, тип шкали, допустимі значення та код пропущеного значення (підрозділ 1.2.1). Файл README описує мету дослідження, вибірку, період збирання даних, структуру пакета, процедури знеособлення та рекомендований спосіб цитування. Текст анкети додають у форматі PDF/A (підрозділ 1.6).

4

Крок 4. Пакування. Файли впорядковують у пакет зі зрозумілою структурою, наприклад: README.txt; data/survey_teachers_2026.csv; docs/codebook.csv; docs/questionnaire.pdf; LICENSE.txt. За потреби пакет описують відповідно до специфікації RO-Crate (підрозділ 1.10.1).

5 **Крок 5. Метадані та ліцензія.** У формі депонування репозитарію заповнюють метадані за схемою DataCite: назву, авторів з ідентифікаторами ORCID iD, установу з ROR ID, анотацію, ключові слова з Тезауруса ЮНЕСКО, рівень освіти за ISCED, тип ресурсу «Dataset», мову, період збирання даних і посилання на пов'язану публікацію (підрозділ 1.4.2). Обирають ліцензію CC BY 4.0 (підрозділ 1.8.1).

6 **Крок 6. Публікація та ідентифікація.** Набір публікують в інституційному репозитарії або в Zenodo (підрозділ 1.7), де йому присвоюється DOI (підрозділ 1.10.2). У статті, що ґрунтується на цих даних, наводять заяву про доступність даних із посиланням на DOI набору (підрозділ 1.10.3).

7 **Крок 7. Перевірка FAIR-відповідності.** Після публікації набір перевіряють сервісом F-UJI за його DOI (підрозділ 1.4.4) і проводять самооцінювання за 15 показниками моделі стандартизації (підрозділ 1.4.3). Наприклад, якщо базові показники (ідентифікатор, метадані, протокол доступу, ліцензія) та більшість інших показників оцінено у 2 бали, а машинну видимість і використання контрольованих словників – у 1 бал, загальний результат може становити близько 22 балів, що відповідає достатньому рівню (Suff-FAIR). Виявлені недоліки (наприклад, відсутність кваліфікованих посилань на пов'язані ресурси) усувають у наступній версії набору, яка отримує власний DOI версії.

Наведений приклад показує, що створення FAIR-пакета не потребує складних технічних засобів: більшість кроків виконується за допомогою табличного й текстового редакторів та форми депонування репозитарію. Визначальними є планування до початку збирання даних і послідовне документування кожного етапу.

2.5. Цифрова компетентність дослідника та інституційне впровадження



Розглянуті в підрозділах 2.1–2.4 практики пошуку, інтеграції, аналізу та застосування FAIR-даних реалізуються не автоматично, а завдяки конкретним компетентностям дослідників і організаційній підтримці наукової установи і ЗВО. Цей завершальний підрозділ розділу об'єднує індивідуальний і інституційний виміри переходу від декларативного визнання принципів FAIR до їх системної практики.

2.5.1. Складники цифрової компетентності наукового працівника

Цифрова компетентність з управління дослідницькими даними охоплює чотири взаємопов'язані складники: технічний (володіння форматами, стандартами метаданих і репозитаріями, розглянутими в розділі 1 довідника), методологічний (уміння оцінити придатність чужого масиву даних до повторного використання за критеріями належної перевірки, розглянутої в підрозділі 2.1.2), правовий та етичний (розуміння меж інформованої згоди й вимог захисту персональних даних, розглянутих у підрозділі 1.9) і аналітичний (застосування методів вторинного аналізу й освітньої аналітики, розглянутих у підрозділі 2.3). Рамка компетентностей FAIR для вищої освіти (FAIR4HE), розглянута в підрозділі 1.1.5, деталізує ці складники для трьох рівнів підготовки дослідника – бакалаврського, магістерського та докторського рівня, визначаючи для кожного з них очікуваний рівень самостійності у плануванні, документуванні та повторному використанні даних [1].

Європейські документи з розвитку компетентностей для відкритої науки розглядають уміння працювати з FAIR-даними як складник ширшої цифрової компетентності дослідника. Звіт робочої групи з навичок і навчання Ради Європейської хмари відкритої науки (EOSC) визначає, що такі компетентності потрібні не лише дослідникам, а й фахівцям, які їх підтримують, – розпорядникам даних, бібліотекарям, IT-фахівцям, – і пропонує рамку для їх розвитку на індивідуальному та інституційному рівнях [2]. Для наукових

установ це означає, що розвиток компетентностей з управління даними має бути спільним завданням науковців і служб підтримки досліджень.

У практичному вимірі цифрову компетентність наукового працівника з управління FAIR-даними можна описати через перелік конкретних умінь. Дослідник уміє скласти план управління даними (підрозділ 1.5); обрати відкритий формат і впорядкувати структуру файлів (підрозділи 1.6, 1.10.1); описати набір даних метаданими й підготувати кодову книгу та README (підрозділи 1.2, 1.4); визначити режим доступу й ліцензію з урахуванням етичних і правових вимог (підрозділи 1.8, 1.9); опублікувати набір даних у репозитарії з присвоєнням DOI (підрозділ 1.10); знайти, оцінити й правомірно використати чужі дані (підрозділи 2.1, 2.2); коректно процитувати використані дані (підрозділ 2.2.3).

Для самооцінювання рівня обізнаності з принципами FAIR зручно використовувати безкоштовний онлайн-інструмент FAIR-Aware: дослідник відповідає на запитання про свій набір даних, а інструмент пояснює, які дії допоможуть зробити дані відповіднішими принципам FAIR, і пропонує посилання на додаткові матеріали [3]. Проходження такого самооцінювання перед публікацією першого набору даних допомагає виявити прогалини в знаннях і визначити напрями подальшого навчання. На рівні установи результати подібного самооцінювання можуть слугувати підставою для планування заходів із підвищення кваліфікації наукових і науково-педагогічних працівників.



Важливо, що цифрова компетентність з управління FAIR-даними не є статичною: стандарти, сервіси та вимоги фінансування журналів постійно змінюються. Тому дослідникові варто розглядати її розвиток як частину безперервного професійного розвитку, періодично оновлюючи знання через навчальні ресурси, розглянуті в підрозділі 2.5.4, та обмін досвідом з колегами в межах установи.

2.5.2. Відповідальне управління даними: етика та академічна доброчесність

Відповідальне управління даними виходить за межі формального дотримання вимог репозитарію чи законодавства про захист персональних даних (підрозділ 1.9) і охоплює ширший вимір академічної доброчесності: коректне визнання внеску авторів первинних масивів даних через цитування з постійним ідентифікатором (підрозділ 2.2.3), прозоре розкриття меж застосовності повторно використаних даних до нового дослідницького

контексту та утримання від вибіркового представлення результатів вторинного аналізу (підрозділ 2.3.1), що спотворює висновки первинного дослідження. Особливого значення ця вимога набуває для здобувачів наукового ступеня, які вперше здійснюють вторинний аналіз чужих даних: формування навички відрізняти повторне використання даних від їх привласнення без належного визнання є складником докторської підготовки, а не лише технічним аспектом цитування.

Європейський кодекс поведінки для доброчесності досліджень, оновлений Федерацією європейських академій наук (ALLEA) у 2023 р., окремо розглядає належні практики роботи з даними: дослідники та установи мають забезпечувати належне управління даними й дослідницькими матеріалами, їх зберігання та доступність «настільки відкриту, наскільки можливо, і настільки закриту, наскільки необхідно», зокрема відповідно до принципів FAIR [4]. Таким чином, управління даними розглядається не як технічне питання, а як невід'ємна частина доброчесної дослідницької практики.

В Україні основні вимоги до академічної доброчесності визначає Закон «Про академічну доброчесність», ухвалений 18 грудня 2025 р. [5]. Серед видів порушень, визначених законом, безпосередньо стосуються роботи з даними фабрикація (вигадування даних чи фактів, що використовуються в дослідженні) та фальсифікація (свідома зміна чи модифікація наявних даних), а також відчуження авторства, приписування авторства й недоброчесне використання результатів, згенерованих штучним інтелектом. Принципи FAIR є дієвим засобом запобігання таким порушенням: якщо первинні дані, кодова книга та скрипти аналізу доступні для перевірки, фабрикацію чи фальсифікацію значно складніше здійснити й легше виявити.

Важливим аспектом доброчесності є також справедливе визнання внеску всіх учасників створення даних. Таксономія ролей учасників дослідження CRediT виокремлює, зокрема, роль «Data curation» (курування даних) – діяльність з опису, очищення та підтримання даних для подальшого використання [6]. Зазначення цієї ролі в публікаціях і метаданих набору даних дає змогу визнати внесок аспірантів, лаборантів чи розпорядників даних, які нерідко виконують значну частину роботи з підготовки даних, але не завжди потрапляють до переліку авторів.

Якщо після публікації набору даних виявлено помилку, дослідник має оприлюднити виправлену версію з поясненням змін (підрозділ 1.7.1), а за потреби повідомити редакцію журналу, у якому опубліковано результати.

Відкрите визнання й виправлення помилок є ознакою доброчесності, а не приводом для осуду.

Для наукових установ відповідальне управління даними означає також створення умов, за яких доброчесна поведінка стає звичною практикою: наявність інституційної політики управління даними, доступ до консультацій розпорядника даних, визнання опублікованих наборів даних як наукових результатів під час оцінювання діяльності працівників (підрозділ 2.5.5).

2.5.3. Підготовка нових дослідників: навчання на реальних масивах даних

Найефективнішою формою формування компетентностей з управління FAIR-даними визнано практичну роботу здобувачів наукового ступеня з реальними, а не навчально-змодельованими масивами даних: аналіз відкритих освітніх мікроданих TIMSS, PIRLS чи PISA (підрозділ 1.7.2) у межах курсової чи дисертаційної роботи дає змогу одночасно опанувати технічні навички роботи з форматами й документацією (підрозділ 1.6) та критичне осмислення обмежень чужих даних (підрозділ 2.1.2). Підготовка здобувачів наукового ступеня в установах і ЗВО України до самостійного формування й використання FAIR-даних відповідно до розглянутих у розділі 1 довідника вимог є практичною реалізацією компетентнісного складника рамки FAIR4HE для докторського рівня підготовки [1].

Аналіз програм підготовки фахівців з управління даними свідчить, що найкращих результатів досягають курси, які поєднують коротку теоретичну частину з практичними завданнями на реальних даних і проєктною роботою, наближеною до професійної діяльності [7]. Для аспірантів у галузі освітніх наук така підготовка може бути побудована як низка послідовних завдань: скласти план управління даними для власного дисертаційного дослідження (підрозділ 1.5.2); підготувати кодову книгу й README до пробного набору даних; опублікувати його в «пісочниці» репозитарію чи як закритий запис; знайти й оцінити придатність відкритого набору даних з теми дослідження (підрозділ 2.1); відтворити один із результатів опублікованої статті на відкритих даних.

Зручним матеріалом для таких завдань є дані міжнародних досліджень якості освіти. Україна вперше взяла участь у дослідженні PISA у 2018 р., а у 2022 р. – вдруге [8], тож аспіранти можуть працювати з даними, які стосуються саме української освіти, порівнювати їх з даними інших країн і на практиці опановувати роботу з документацією, вибірковими вагами й правдоподібними значеннями (підрозділ 2.3.1). Повні масиви даних і технічна документація PISA доступні на сайті OECP [9].

Для початкового опанування навичок упорядкування даних корисними є відкриті навчальні матеріали спільноти The Carpentries, зокрема урок «Data Organization in Spreadsheets for Social Scientists», який на простих прикладах пояснює, як організувати дані в електронних таблицях, уникати типових помилок форматування та експортувати дані у відкриті формати [10]. Такі матеріали можна використовувати як основу для практичних занять в аспірантурі, адаптувавши приклади до тематики педагогічних досліджень.



Важливо, щоб навчання не обмежувалося технічними навичками. Аспіранти мають обговорювати етичні кейси (наприклад, чи можна оприлюднити відеозапис уроку, чи охоплює згода батьків повторне використання даних), аналізувати приклади вдалих і невдалих заяв про доступність даних, оцінювати якість документації чужих наборів даних. Науковий керівник, своєю чергою, може підтримати формування цих навичок, обговорюючи з аспірантом план управління даними на початку дослідження і перевіряючи підготовку пакета даних перед захистом дисертації.

Корисно, щоб кожен аспірант до завершення підготовки мав щонайменше один опублікований набір даних із DOI, пов’язаний із його дисертаційним дослідженням, і профіль ORCID із зазначеними публікаціями та даними. Це стає не лише свідченням сформованих компетентностей, а й складником наукового портфоліо молодого дослідника.

2.5.4. Ресурси й освітні платформи (The Turing Way, FOSTER+, OpenAIRE, Рамка компетентностей FAIR4HE)

Для самостійного опанування практик формування й використання FAIR-даних дослідникові доступна низка відкритих освітніх ресурсів. The Turing Way – спільнотний, відкрито створюваний посібник із відтворюваних, етичних і спільних досліджень, що містить практичні розділи з управління дослідницькими даними та застосування принципів FAIR, розроблюваний і підтримуваний Інститутом Алана Тюрінга (Велика Британія) [11]. FOSTER+ (Facilitate Open Science Training for European Research) – платформа електронного навчання з відкритої науки, що пропонує структуровані навчальні курси, безпосередньо присвячені принципам FAIR і плануванню управління даними (підрозділ 1.5), орієнтовані на дослідників, бібліотекарів та адміністраторів наукових проєктів. OpenAIRE, згаданий у підрозділах 1.5.1, 1.9.2 та 2.1.1,окрім інфраструктурної функції, підтримує серію навчальних

вебінарів і настанов із питань відкритої науки та управління даними, адаптованих до вимог програми Horizon Europe. Рамка компетентностей FAIR4HE [1] слугує організаційною основою для побудови власної навчальної програми установи на основі цих ресурсів, дозволяючи зіставити наявні курси з конкретними компетентностями, визначеними для кожного рівня підготовки дослідника.

Портал FOSTER Open Science продовжує надавати доступ до навчальних матеріалів і курсів, створених у межах проєктів FOSTER і FOSTER+, зокрема з ліцензування результатів досліджень, відкритого рецензування та управління даними [12]. Європейський проєкт Skills4EOSC, що виконувався в межах програми «Горизонт Європа» у 2022–2025 рр., спрямований на створення мережі центрів компетентностей з відкритої науки та спільної системи навчання й сертифікації фахівців, здатних працювати з FAIR-даними [13]. Корисними є також матеріали, розроблені архівами даних соціальних наук, зокрема Посібник експерта з управління даними консорціуму CESSDA, побудований за етапами життєвого циклу даних і орієнтований саме на дослідників соціальних і поведінкових наук, до яких належать і освітні дослідження [14].

Більшість зазначених ресурсів доступні англійською мовою, тож для українських освітян доцільно поєднувати їх із вітчизняними матеріалами: методичними рекомендаціями Міністерства освіти і науки України (МОН України) щодо управління науковими даними [15], публікаціями співробітників ІЦО НАПН України з питань FAIR-даних, зокрема щодо понятєво-термінологічного апарату [16], таксономії [17] та огляду цифрових сервісів [18], а також цим довідником.

2.5.5. Інституційне впровадження: політики, інфраструктура, бар’єри, показники результативності (KPI)

Індивідуальна компетентність дослідника реалізується повною мірою лише за наявності інституційної підтримки: політики управління даними, технічної інфраструктури (підрозділ 1.7) та служб, що супроводжують дослідника на кожному етапі життєвого циклу даних (підрозділ 1.3.1). Систематизація підходів до оцінювання готовності наукової установи до впровадження таких служб представлена в моделі зрілості організаційних служб управління дослідницькими даними, опублікованій у 2025 р.: модель виокремлює шість вимірів: обізнаність (взаємодія між структурними підрозділами), політика управління даними, бюджетне забезпечення, спектр послуг, розуміння потреб

користувачів та ІТ-інфраструктура, кожен з яких оцінюється за п’ятирівневою шкалою зрілості (табл. 2.2) від початкового (безсистемні, ситуативні дії) до оптимізованого (регулярний перегляд і стратегічне вдосконалення на основі зворотного зв’язку) [19].

Таблиця 2.2.

Рівні інституційної зрілості управління дослідницькими даними

Рівень	Характеристика	Як визначити на практиці (орієнтовні індикатори)
Початковий	Дії з управління даними безсистемні, здійснюються на розсуд окремих підрозділів чи дослідників без узгодженої політики	Політику управління даними не затверджено; план управління даними складають в окремих випадках (менше ніж для 10 % НДР); набори даних публікують епізодично, переважно без DOI; відповідальних за управління даними не визначено
Відтворюваний	Основні напрями й процедури визначено, керовані на рівні окремого підрозділу	Діють рекомендації чи процедури окремих підрозділів; плани управління даними мають 10–50 % НДР; частина наборів даних має DOI; у підрозділах визначено відповідальних осіб; навчання дослідників проводиться епізодично
Визначений	Критерії оцінювання для кожного виміру задокументовано й узгоджено на рівні установи	Затверджено інституційну політику управління даними; план управління даними є обов’язковим для нових НДР; діє репозитарій з присвоєнням DOI; визначено розпорядника даних установи; критерії оцінювання FAIR-відповідності (підрозділ 1.4.3) задокументовано
Керований	Цільові показники за кожним виміром активно відстежуються, прогрес піддається кількісному вимірюванню керівництвом установи	Щорічно збирають і подають керівництву показники: частку НДР із планом управління даними (понад 90 %), частку наборів даних із DOI, середній бал FAIR-відповідності; більшість наборів досягає щонайменше базового рівня (Min-FAIR); навчання дослідників систематичне
Оптимізований	Процедури оцінювання регулярно переглядаються; керівництво завчасно враховує виклики й здійснює стратегічні коригування	Показники та цільові значення переглядають щороку; більшість наборів даних досягає достатнього (Suff-FAIR) і вищого рівня; установа інтегрована з національною та європейською інфраструктурою відкритої науки (EOSC) і поширює власний досвід

Складено за [19] із доповненням авторів.

У межах вимірів «спектр послуг» та «IT-інфраструктура» модель конкретизує ключові функційні області, придатні до використання як показники результативності (KPI) для установ НАПН України: складання планів управління даними (підрозділ 1.5), навчання дослідників з інформаційної грамотності (підрозділ 2.5.3), управління правами інтелектуальної власності (підрозділ 1.8), зберігання даних, публікація даних, курування масивів даних та підтримка отримання й цитування даних (підрозділ 2.2.3) для виміру послуг; а також сховище даних, репозитарій даних, платформа аналізу даних та інформаційна система досліджень – для виміру інфраструктури [19].

Основними бар'єрами інституційного впровадження, задокументованими в емпіричних дослідженнях 2025 р., є не технічна складність, а організаційна: розпорядники даних (data steward, підрозділ 1.3.2) часто змушені самостійно інтерпретувати загальні вимоги інституційної політики стосовно конкретних ситуацій, особливо щодо рівня чутливості даних і меж анонімізації (підрозділ 1.9.2), що призводить до розбіжностей у застосуванні політики між структурними підрозділами й зумовлює потребу в регулярній міжпідрозділовій координації [20]. Досвід застосування моделі зрілості до національної вибірки установ засвідчив, що рівень розвитку IT-інфраструктури помітно корелює з розміром установи та наявністю зовнішніх нормативних вимог, тоді як вимір бюджетного забезпечення – найповільніший у розвитку компонент, що вказує на потребу цілеспрямованого фінансового планування для установ, які лише розпочинають розбудову інституційної інфраструктури FAIR-даних, розглянутої в підрозділі 1.7.4 [19].

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ДО РОЗДІЛУ 2



Література до 2.1.

1. Іванова С. М., Кільченко А. В. Генеза понятійного концепту з формування та використання FAIR-даних у галузі освітніх наук. *Освіта. Інноватика. Практика.* 2026. Вип. 14 (4). С. 28–35. DOI: https://doi.org/10.31110/2616-650X-vol14i4-004
2. Gregory K. A dataset describing data discovery and reuse practices in research. *Scientific Data.* 2020. Vol. 7. Article 232. DOI: https://doi.org/10.1038/s41597-020-0569-5
3. Yoon A. Data reusers' trust development. *Journal of the Association for Information Science and Technology*. 2017. Vol. 68. Iss. 4. P. 946–956. DOI: https://doi.org/10.1002/asi.23730
4. Re(Use) Data Guide: Data Reuse Checklist. *Yale University, Cushing/Whitney Medical Library*, 2025. URL: https://library.medicine.yale.edu/wp-content/uploads/2025/01/data_reuse_checklist.pdf

Література до 2.2.

1. Metadata Vocabulary for Tabular Data / J. Tennison, G. Kellogg (eds.). *W3C Recommendation,* 17 December 2015. URL: https://www.w3.org/TR/tabular-metadata/
2. Pasquetto I. V., Randles B. M., Borgman C. L. On the Reuse of Scientific Data. *Data Science Journal.* 2017. Vol. 16. Article 8. DOI: https://doi.org/10.5334/dsj-2017-008
3. Yoon A. Data reusers' trust development. *Journal of the Association for Information Science and Technology.* 2017. Vol. 68. Iss. 4. P. 946–956. DOI: https://doi.org/10.1002/asi.23730
4. Gerasimov I., Savtchenko A., Alfred J., Acker J., Wei J., KC B. Bridging the Gap: Enhancing Prominence and Provenance of NASA Datasets in Research Publications. *Data Science Journal*. 2024. Vol. 23. Article 1. DOI: https://doi.org/10.5334/dsj-2024-001
5. Kedron P., Li Z., Liu L. The computational reproducibility of articles published under the Open Data + FAIR policy of IJGIS. *International Journal of Geographical Information Science*. 2026. P. 1–22. DOI: https://doi.org/10.1080/13658816.2025.2603586
6. Data Citation Synthesis Group: Joint Declaration of Data Citation Principles / M. Martone (ed.). *San Diego, CA: FORCE11*, 2014. DOI: https://doi.org/10.25490/a97f-egyk

Література до 2.3

1. Іванова С. М., Кільченко А. В. Генеза понятійного концепту з формування та використання FAIR-даних у галузі освітніх наук. *Освіта. Інноватика. Практика.* 2026. Вип. 14 (4). С. 28–35. DOI: https://doi.org/10.31110/2616-650X-vol14i4-004
2. Mao Z. AI-related measures and self-reported lesson-aim attainment: an explainable machine-learning analysis of TALIS 2024. Frontiers in Public Health. 2026. Vol. 14. Article 1911565. DOI: https://doi.org/10.3389/fpubh.2026.1911565
3. Rutkowski L., Gonzalez E., Joncas M., von Davier M. International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting. *Educational Researcher.* 2010. Vol. 39. Iss. 2. P. 142–151. DOI: https://doi.org/10.3102/0013189X10363170
4. IEA. Data Tools: IEA IDB Analyzer. *International Association for the Evaluation of Educational Achievement.* URL: https://www.iea.nl/data-tools/tools
5. Siemens G., Long P. Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review. 2011. Vol. 46. № 5. P. 30–40.
6. Romero C., Ventura S. Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery.* 2020. Vol. 10. Iss. 3. Article e1355. DOI: https://doi.org/10.1002/widm.1355
7. Jamovi – open statistical software. URL: https://www.jamovi.org
8. Pellegrini M., Day E., Scarbrough H. F., Pigott T. D. A Meta-Review of Education Meta-Analyses: Relevance, Applicability, and Accessibility of Findings. AERA Open. 2025. Vol. 11. DOI: https://doi.org/10.1177/23328584251389562
9. Page M. J., McKenzie J. E., Bossuyt P. M. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021. Vol. 372. Article n71. DOI: https://doi.org/10.1136/bmj.n71
10. OECD. Everybody Cares About Using Education Research Sometimes: Perspectives of Knowledge Intermediaries. Educational Research and Innovation. *Paris: OECD Publishing,* 2025. 170 p. DOI: https://doi.org/10.1787/5ef88972-en
11. Education Endowment Foundation. Teaching and Learning Toolkit. URL: https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit
12. Slavin R. E. How evidence-based reform will transform research and practice in education. Educational Psychologist. 2020. Vol. 55. Iss. 1. P. 21–31. DOI: https://doi.org/10.1080/00461520.2019.1611432

Література до 2.4.

1. Data for Psychological Research in the Educational Field: Spotlights, Data Infrastructures, and Findings from Research / S. Bayer, K. Blask, T. Gnambs et al. *Journal of Open Psychology Data*. 2023. Vol. 1. Iss. 1. Article 19. DOI: https://doi.org/10.5334/jopd.105
2. Bittermann A., Lauer T., Löhlein A. M., Omieczynski C., Gollwitzer M., Sassenberg K. The Status Quo of FAIR Data Sharing in Psychology. *Preprint. PsychArchives,* 2025. URL: https://www.psycharchives.org/en/item/a16acd12-b43c-4b26-9134-e7706e702e91

3. Liu Q., Khalil M. Understanding privacy and data protection issues in learning analytics using a systematic review. *British Journal of Educational Technology*. 2023. Vol. 54. Iss. 6. P. 1715–1747. DOI: https://doi.org/10.1111/bjet.13388

4. Drachsler H., Greller W. Privacy and analytics: it's a DELICATE issue. A checklist for trusted learning analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16)*. New York: ACM, 2016. P. 89–98. DOI: https://doi.org/10.1145/2883851.2883893

5. Tsai Y. S., Moreno-Marcos P. M., Jivet I., Scheffel M., Tammets K., Kollom K., Gašević D. The SHEILA Framework: Informing Institutional Strategies and Policy Processes of Learning Analytics. *Journal of Learning Analytics.* 2018. Vol. 5. Iss. 3. P. 5–20. DOI: https://doi.org/10.18608/jla.2018.53.2

6. Liu Q., Khalil M. Understanding privacy and data protection issues in learning analytics using a systematic review. British Journal of Educational Technology. 2023. Vol. 54. Iss. 6. P. 1715–1747. DOI: https://doi.org/10.1111/bjet.13388

7. Liu Q., Khalil M., Jovanovic J., Shakya R. Scaling while privacy preserving: a comprehensive synthetic tabular data generation and evaluation in learning analytics. *Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK '24). New York: ACM,* 2024. P. 620–631. DOI: https://doi.org/10.1145/3636555.3636868

8. A privacy preserving synthetic learner dataset for learning analytics in technology enhanced higher education. Scientific Reports. 2026. DOI: https://doi.org/10.1038/s41598-026-44990-8

Література до 2.5.

1. FAIR Competence Framework for Higher Education (Data Stewardship Professional Competence Framework) / Y. Demchenko et al. *FAIRsFAIR Project Deliverable D7.3. Zenodo,* 2021. DOI: https://doi.org/10.5281/zenodo.4562089

2. Digital skills for FAIR and Open Science: Report from the EOSC Executive Board Skills and Training Working Group. *Luxembourg: Publications Office of the European Union*, 2021. DOI: https://doi.org/10.2777/59065

3. FAIR-Aware: Assess Your Knowledge of FAIR (v1.0.1) / V. Akerman, L. Cepinskas, M. Verburg, M. Mustapha. *Zenodo,* 2021. DOI: https://doi.org/10.5281/zenodo.5084861

4. ALLEA. The European Code of Conduct for Research Integrity. Revised Edition 2023. Berlin: ALLEA, 2023. URL: https://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf

5. Про академічну доброчесність: Закон України від 18.12.2025 № 4742-IX. URL: https://zakon.rada.gov.ua/laws/show/4742-20

6. CRediT – Contributor Roles Taxonomy. ANSI/NISO Z39.104-2022. *National Information Standards Organization (NISO).* URL: https://credit.niso.org

7. Oladipo F., Folorunso S., Ogundepo E., Osigwe O., Akindele A. Curriculum Development for FAIR Data Stewardship. *Data Intelligence*. 2022. Vol. 4. Iss. 4. P. 991–1012. DOI: https://doi.org/10.1162/dint_a_00183

8. Міжнародне дослідження PISA-2022 в Україні. *Освіта.ua.* URL: https://osvita.ua/school/90711

9. OECD. PISA Data. URL: https://www.oecd.org/en/about/programmes/pisa/pisa-data.html

10. Data Carpentry. Data Organization in Spreadsheets for Social Scientists. *The Carpentries.* URL: https://datacarpentry.github.io/spreadsheets-socialsci/

11. The Turing Way Community. The Turing Way: A Handbook for Reproducible, Ethical and Collaborative Research. *The Alan Turing Institute.* URL: https://the-turing-way.netlify.app

12. FOSTER Open Science: e-learning portal. URL: https://www.fosteropenscience.eu

13. Skills4EOSC: Skills for the European Open Science Commons. Horizon Europe project (GA 101058527). URL: https://www.skills4eosc.eu

14. CESSDA Training Team. CESSDA Data Management Expert Guide. *Bergen: CESSDA ERIC*, 2020. DOI: https://doi.org/10.5281/zenodo.3820473

15. Методичні рекомендації щодо управління науковими даними для закладів вищої освіти та наукових установ / Міністерство освіти і науки України. Київ, 2024. URL: https://mon.gov.ua/static-objects/mon/sites/1/nauka/2024/12/31/metod-rekomendatsiyi-shchodo-upravlinnya-naukovymy-danymy-31-12-2024.pdf

16. Іванова С. М., Кільченко А. В. Генеза понятійного концепту з формування та використання FAIR-даних у галузі освітніх наук. *Освіта. Інноватика. Практика.* 2026. Вип. 14 (4). С. 28–35. DOI: https://doi.org/10.31110/2616-650X-vol14i4-004

17. Осадча К. П., Спірін О. М., Олексюк В. П. Таксономія FAIR-даних для наукових досліджень у галузі освіти в умовах цифрової трансформації. *Вісник Глухівського НПУ ім. О. Довженка*. 2026. Вип. 2 (61). С. 10–22. DOI: https://doi.org/10.31376/2410-0897-2026-2-61-10-22

18. Вакалюк Т. А., Мінтій І. С. Огляд наявних цифрових сервісів для зберігання й опрацювання FAIR даних в галузі освітніх наук. *Наукові записки Бердянського державного педагогічного університету. Педагогічні науки*. 2026. № 1. С. 357–368. DOI: https://doi.org/10.32782/2412-9208-2026-1-357-368

19. Minamiyama Y., Ikeuchi U., Tanabe K., Yuki K., Hayashi K., Aoki T. Maturity Model for Organizational Research Data Management Services. *Data Science Journal*. 2025. Vol. 24. Article 18. DOI: https://doi.org/10.5334/dsj-2025-018

20. Stepanovic O. et al. Open research data integration in universities: How data stewards adapt global policies to local contexts. *Journal of the Association for Information Science and Technology.* 2025. DOI: https://doi.org/10.1002/asi.70034

ВИСНОВКИ



Стрімке зростання обсягів дослідницьких даних у галузі освітніх наук, розглянуте в передмові до цього довідника, супроводжується не менш стрімким ускладненням вимог до їх упорядкованого зберігання, опису та повторного використання. Довідник підготовлено як науково-методичний результат першого етапу виконання НДР «Технології формування і використання FAIR даних в галузі освітніх наук» (ДР № 0126U000521) і присвячено систематизації теоретичних засад та практичного інструментарію застосування принципів FAIR (Findable, Accessible, Interoperable, Reusable) до дослідницьких даних педагогічної науки.

У розділі І довідника викладено цілісний цикл формування FAIR-даних. Розкрито зміст і еволюцію концепції FAIR та кожного з чотирьох принципів – відшукуваності, доступності, сумісності й багаторазовості – з наголосом на тому, що FAIR не тотожне безумовній відкритості: дані з обмеженим доступом також можуть відповідати принципам FAIR, якщо вони належно описані, мають постійний ідентифікатор і чітко визначені умови доступу. Запропоновано багатовимірну типологію освітніх наукових даних, що враховує їх природу (кількісні, якісні, змішані), походження (первинні дослідницькі, адміністративні та інституційні дані), а також тип і метод дослідження, у межах якого їх отримано; показано, що саме тип даних визначає вимоги до їх документування (кодова книга для кількісних даних, протокол аналізу для якісних), вибору формату й режиму доступу. Розглянуто життєвий цикл дослідницьких даних – від планування до архівування чи контрольованого видалення – і розподіл відповідальності між керівником дослідження, розпорядником даних, фахівцем із захисту персональних даних та IT-супроводом.

Центральним теоретичним результатом розділу 1 є модель стандартизації FAIR-даних для галузі освітніх наук, що складається з цільового, змістового, процесуального й результативного блоків (рис. 1.1). Результативний блок містить п’ятнадцять показників, згрупованих за чотирма критеріями F, A, I, R,

кожен з яких оцінюється від 0 до 2 балів (максимум – 30 балів), і шкалу чотирьох рівнів відповідності: недостатнього (0–10 балів), базового (11–18), достатнього (19–25) і високого (26–30) (табл. 1.1). Визначено чотири базові показники – постійний ідентифікатор, метадані, стандартизований протокол доступу та ліцензію, – без яких набір даних не може досягти навіть базового рівня. Модель перетворює загальні принципи FAIR на перелік конкретних запитань для самоперевірки й дає змогу перейти від декларативного визнання відповідності набору даних принципам FAIR до його відтворюваного кількісного оцінювання, а її зіставлення з моделлю FAIR-зрілості RDA та інструментом F-UJI засвідчило узгодженість із міжнародними підходами за збереження галузевої специфіки (контрольовані словники ISCED, стандарти xAPI та IEEE LOM).

Практично зорієнтовані підрозділи розділу 1 містять інструменти, які дослідник може безпосередньо застосувати у власній роботі: заповнюваний шаблон плану управління даними (табл. 1.2), узгоджений із базовими вимогами Science Europe та Методичними рекомендаціями МОН України; рекомендації щодо відкритих форматів зберігання для різних типів освітніх даних (табл. 1.3); порівняльну характеристику загальнонаукових репозитаріїв Zenodo, Figshare, OSF і Dryad (табл. 1.4) та платформ інституційних репозитаріїв (табл. 1.5); огляд національної інфраструктури – НРАТ і Електронної бібліотеки НАПН України, – захищених архівів для чутливих даних (на прикладі FAIRVault) та сервісів Європейської хмари відкритої науки (EOSC). Розглянуто вибір відкритих ліцензій Creative Commons та Open Data Commons (табл. 1.6), процедури автентифікації й авторизації доступу, нормативну базу захисту персональних даних (GDPR і Закон України «Про захист персональних даних»), методи та безкоштовні інструменти анонімізації й псевдонімізації (табл. 1.7), а також завершальні кроки публікації: пакування дослідницьких об’єктів (RO-Crate, BagIt), присвоєння постійних ідентифікаторів (DOI, ORCID, ROR), підготовку заяви про доступність даних, завантаження та перевірку масиву в репозитарії.

У розділі II довідника розглянуто процес використання вже оприлюднених FAIR-даних іншими дослідниками. Описано стратегії пошуку наукових даних за допомогою Google Dataset Search, DataCite Search, BASE та OpenAIRE Explore і процедуру належної перевірки масиву перед повторним використанням: контрольний перелік (табл. 2.1) допомагає оцінити ліцензію, обсяг інформованої згоди учасників, ризик повторної ідентифікації, повноту документування, походження й метод збору, технічну сумісність, цілісність і

версію даних, а також правила їх цитування. Окремо висвітлено правові та етичні аспекти доступу до масивів з обмеженим режимом. Показано, що поєднання даних з різних джерел потребує не лише технічного узгодження форматів, а й семантичного узгодження термінології за допомогою спільних класифікаторів (ISCED, EuroVoc), а коректне цитування набору даних з постійним ідентифікатором і документування його походження є умовою відтворюваності дослідження.

Охарактеризовано основні напрями аналітичного застосування FAIR-даних: вторинний аналіз, зокрема методами машинного навчання; кількісний і якісний аналіз із використанням доступних освітянам безкоштовних інструментів; метааналіз і систематичні огляди; міжнародні порівняльні дослідження на основі мікроданих TIMSS, PIRLS і PISA; розроблення доказової освітньої політики. Розглянуті приклади психолого-педагогічних опитувань, навчальної аналітики та синтетичних даних, а також покрокова процедура створення FAIR-пакета, засвідчили, що найбільші ризики в освітніх дослідженнях пов’язані не стільки з технічними, скільки з етичними питаннями – захистом персональних даних учасників освітнього процесу, зокрема неповнолітніх, і дотриманням меж їхньої інформованої згоди. Завершує розділ виклад складників цифрової компетентності дослідника, засад відповідального управління даними й академічної доброчесності, підходів до підготовки нових дослідників на реальних масивах даних, навчальних ресурсів (The Turing Way, FOSTER+, OpenAIRE, рамка компетентностей FAIR4HE) та умов інституційного впровадження, зокрема рівнів інституційної зрілості управління дослідницькими даними (табл. 2.2) і показників результативності для установ НАПН України.

Наскрізною ідеєю довідника є положення про те, що формування та використання FAIR-даних є не двома ізольованими технічними процедурами, а єдиним циклом наукової комунікації: якість рішень, ухвалених на етапі планування й документування дослідження, безпосередньо визначає науковий і практичний потенціал даних для їхнього подальшого застосування іншими дослідниками – вторинного аналізу, перевірки результатів, метааналітичних узагальнень і формування доказової освітньої політики. Водночас проведений аналіз засвідчив, що специфіка освітніх наукових даних – їх соціальна чутливість, різноманітність форматів і необхідність дотримання етичних та правових вимог щодо учасників освітнього процесу, зокрема неповнолітніх, – потребує послідовної адаптації загальнонаукових підходів до управління

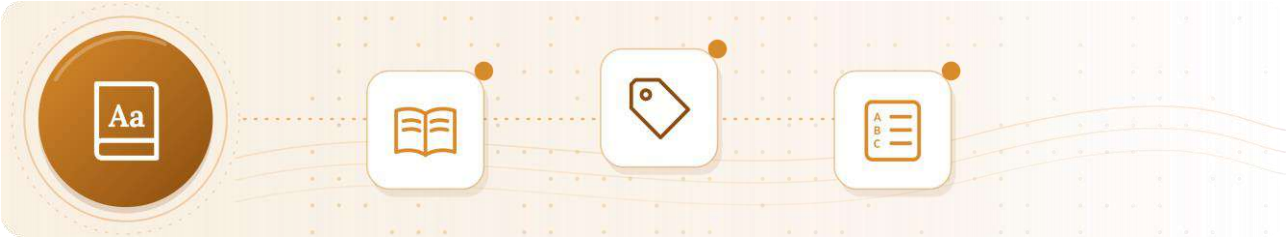
даними, розроблених передусім у природничих і технічних науках, до реалій педагогічних досліджень.

Практичне значення довідника для наукових і науково-педагогічних працівників полягає в наданні узгодженого, вимірюваного інструментарію оцінювання FAIR-відповідності власних дослідницьких даних, а також в окресленні конкретних кроків розвитку інституційної інфраструктури – від модернізації наявних репозитаріїв до розвитку компетентності з управління дослідницькими даними відповідно до вимог сучасної відкритої науки та європейського дослідницького простору.

За результатами узагальнення матеріалів довідника можна сформулювати низку практичних рекомендацій для наукових і науково-педагогічних працівників. *По-перше,* управління даними варто планувати ще до початку їх збирання: складати план управління даними та узгоджувати з учасниками дослідження обсяг інформованої згоди на подальше використання даних. *По-друге*, кожен масив даних має супроводжуватися кодовою книгою або протоколом аналізу та файлом README і зберігатися у відкритих форматах. *По-третє*, для оприлюднення доцільно обирати надійні репозитарії, що присвоюють DOI, і зазначати відкриту ліцензію; дані, що містять персональну інформацію, слід оприлюднювати лише після анонімізації або надавати в режимі обмеженого доступу. *По-четверте*, перед публікацією варто самостійно оцінити набір даних за показниками моделі стандартизації, а використовуючи дані інших дослідників – перевіряти їх за контрольним переліком і коректно цитувати. *По-п'яте*, наукові установи мають підтримувати дослідників організаційно: ухвалювати політики управління даними, визначати розпорядників даних, розвивати інституційні репозитарії та враховувати опубліковані набори даних під час оцінювання наукової діяльності.

Розвиток практик формування і використання FAIR-даних, розглянутий у цьому довіднику, є довгостроковим і поступовим процесом, що потребує узгодженої участі дослідників, наукових установ і органів управління освітою, однак саме такий послідовний розвиток є умовою повноцінної інтеграції української педагогічної науки в міжнародний простір відкритої науки.

ГЛОСАРІЙ



- **FAIR-дані в галузі освітніх наук** – сукупність цифрових дослідницьких даних і метаданих, отриманих або використаних у педагогічних/освітніх дослідженнях (кількісних, якісних чи змішаних), які зберігаються та поширюються відповідно до принципів FAIR (Findable, Accessible, Interoperable, Reusable), що забезпечує їхню відшукуваність, доступність, сумісність і повторне використання.
- **Автентифікація** – перевірка того, що користувач, який входить до системи чи репозитарію, справді є тим, за кого себе видає (наприклад, вхід за логіном і паролем або через ORCID iD).
- **Авторизація** – надання користувачеві, особу якого вже перевірено, певних прав: переглядати, завантажувати чи змінювати дані.
- **Агрегатор наукових даних** – пошуковий сервіс, який збирає описи (метадані) наборів даних з багатьох репозитаріїв і дає змогу шукати їх в одному місці, наприклад Google Dataset Search, DataCite Search, BASE, OpenAIRE. Сам агрегатор дані не зберігає, а лише вказує, де їх знайти.
- **Анонімізація** – опрацювання даних, після якого встановити особу учасника дослідження неможливо навіть із залученням додаткової інформації. Наприклад, з результатів анкетування вилучають прізвища, назви закладів освіти та інші відомості, за якими можна впізнати респондента. Анонімізовані дані вже не вважаються персональними.
- **Багаторазовість використання (Reusable; придатність до повторного використання)** – четвертий принцип FAIR: дані мають бути описані настільки докладно (методика, вибірка, інструментарій, походження даних) і мати настільки зрозумілі умови використання (ліцензію), щоб інший дослідник міг коректно використати їх у новому дослідженні.
- **Використання FAIR-даних у галузі освітніх наук** – сукупність методологічно обґрунтованих і етично відповідальних практик пошуку, оцінювання, доступу, інтерпретації та застосування освітніх

дослідницьких даних і метаданих, організованих відповідно до принципів FAIR (Findable, Accessible, Interoperable, Reusable), з метою верифікації наукових результатів, вторинного аналізу, порівняльних і метааналітичних досліджень, розроблення доказової освітньої політики та вдосконалення педагогічних практик із забезпеченням їх повторного використання та інтеграції з іншими джерелами даних.

- **Відкрита наука** – підхід до організації наукової діяльності, за якого публікації, дані, методики, програмне забезпечення та освітні ресурси стають якомога доступнішими для суспільства. Принципи FAIR є одним зі складників відкритої науки.
- **Відкритий доступ (Open Access)** – режим, за якого дані чи публікації можна безоплатно переглядати й завантажувати без спеціального дозволу, дотримуючись умов ліцензії.
- **Відкритий формат** – формат файлу, опис якого загальнодоступний і який можна відкрити різними, зокрема безоплатними, програмами (наприклад, CSV, TXT, ODT, PDF/A). Відкриті формати рекомендовано використовувати для довготривалого зберігання даних.
- **Відкриті дані** – дані, які будь-хто може вільно отримувати, використовувати й поширювати за умови дотримання ліцензії (наприклад, із зазначенням авторства).
- **Відтворюваність дослідження** – можливість для іншого дослідника отримати ті самі результати, використовуючи опубліковані дані, методику та опис їх опрацювання.
- **Відшукуваність (Findable)** – перший принцип FAIR: дані та їхній опис мають легко знаходити як людина, так і пошукові системи завдяки постійному ідентифікатору та якісним метаданим.
- **Вторинний аналіз даних** – дослідження, у якому науковець аналізує не власні, а раніше зібрані іншими дослідниками дані, щоб перевірити їхні висновки або відповісти на нове дослідницьке питання (наприклад, аналіз відкритих даних PISA чи TALIS).
- **Диференційна приватність** – математичний метод захисту приватності, за якого до результатів аналізу додають контрольовані випадкові відхилення, щоб за оприлюдненими статистичними показниками не можна було дізнатися відомості про конкретну особу.

- **Довготривале зберігання** – збереження даних упродовж багатьох років так, щоб вони залишалися цілими, доступними та придатними для відкриття сучасними програмами.
- **Доказова освітня політика (evidence-based policy)** – ухвалення управлінських рішень в освіті на основі перевірених наукових даних і результатів досліджень, а не лише досвіду чи інтуїції.
- **Дослідницькі дані** – усі матеріали, зібрані або створені під час дослідження, на яких ґрунтуються його висновки: результати тестування й анкетування, протоколи спостережень, записи й транскрипти інтерв'ю, журнали активності в цифрових платформах, а також супровідні матеріали (анкети, кодові книги, програмний код опрацювання).
- **Доступність (Accessible)** – другий принцип FAIR: має бути зрозуміло, як отримати дані, а доступ до них має надаватися за прозорими правилами. Доступність не означає обов'язкової відкритості: чутливі освітні дані можуть надаватися лише за запитом, але їхній опис має бути доступним завжди.
- **Дублінське ядро (Dublin Core)** – найпоширеніший простий стандарт метаданих, що містить 15 основних елементів опису ресурсу (назва, автор, тема, опис, дата, формат, мова, права тощо). Застосовується в більшості інституційних репозитаріїв та електронних бібліотек.
- **Ембарго** – встановлений автором період, упродовж якого опубліковані в репозитарії дані залишаються закритими; після його завершення вони автоматично стають доступними. Опис даних (метадані) доступний і під час ембарго.
- **Європейська хмара відкритої науки (European Open Science Cloud, EOSC)** – європейська інфраструктура, що об'єднує репозитарії, сервіси та інструменти для зберігання дослідницьких даних і обміну ними; українські дослідники можуть користуватися нею завдяки асоціації України з програмою «Горизонт Європа».
- **Життєвий цикл дослідницьких даних** – послідовність етапів роботи з даними: планування, збирання й опрацювання, опис і документування, зберігання та публікація, доступ і повторне використання, архівування або знищення.
- **Загальний регламент ЄС про захист даних (General Data Protection Regulation, GDPR)** – нормативний акт ЄС, що визначає правила збирання, опрацювання й захисту персональних даних. Його вимоги

враховують під час міжнародних досліджень і публікації даних у європейських репозитаріях.

- **Закриті дані** – дані, які не надаються стороннім користувачам (наприклад, через ризик ідентифікації учасників дослідження). Навіть у цьому разі рекомендовано оприлюднювати їхній опис і пояснення, чому доступ обмежено.
- **Заява про доступність даних (Data Availability Statement)** – коротке повідомлення у статті чи звіті про те, де зберігаються дані дослідження, як і на яких умовах їх можна отримати (або чому вони недоступні).
- **Знеособлення даних** – загальна назва дій, спрямованих на те, щоб дані не можна було пов'язати з конкретною особою; охоплює анонімізацію та псевдонімізацію.
- **Інформована згода** – добровільна згода учасника дослідження (для неповнолітніх – також батьків або інших законних представників) на участь у ньому та на використання його даних, надана після ознайомлення з метою дослідження, способами зберігання й можливим оприлюдненням даних.
- **Керівник дослідження (Principal Investigator, PI)** – науковець, відповідальний за дослідження загалом, зокрема за складання плану управління даними й ухвалення рішень щодо їх оприлюднення.
- **Кодова книга (codebook)** – документ, що пояснює структуру набору даних: перелік змінних, їх назви, значення кодів (наприклад, 1 – «так», 2 – «ні»), одиниці вимірювання, позначення пропущених відповідей. Без кодової книги інший дослідник не зможе правильно зрозуміти дані.
- **Контрольований (обмежений) доступ** – режим, за якого дані надаються лише після виконання певних умов: подання заявки, підтвердження мети використання, підписання угоди про використання даних.
- **Контрольований словник (тезаурус)** – узгоджений перелік термінів певної галузі з визначеними зв'язками між ними, який використовують для однозначного опису даних. Для освіти це, наприклад, Тезаурус ЮНЕСКО з питань освіти, Тезаурус ERIC, EuroVoc.
- **Ліцензія на дані** – визначені автором умови, на яких інші можуть використовувати, поширювати та змінювати дані. Найпоширенішими є ліцензії Creative Commons, наприклад CC BY 4.0 – дозволено будь-яке використання із зазначенням автора.

- **Машинозчитуваність** – властивість даних і метаданих, завдяки якій комп'ютерні програми можуть автоматично знаходити, читати та опрацьовувати їх без участі людини.
- **Метааналіз** – статистичне узагальнення результатів кількох незалежних досліджень однієї проблеми, що дає змогу отримати надійніший загальний висновок (наприклад, про ефективність певної методики навчання).
- **Метадані** – «дані про дані»: структурований опис набору даних (назва, автори, анотація, ключові слова, дата, методика збирання, рівень освіти, умови доступу, ліцензія тощо), який допомагає знайти, зрозуміти й правильно використати дані.
- **Міжнародна стандартна класифікація освіти (International Standard Classification of Education, ISCED)** – розроблена ЮНЕСКО класифікація рівнів освіти від 0 (освіта дітей раннього віку та дошкільна освіта) до 8 (рівень доктора філософії або еквівалентний). Її використання в описі даних забезпечує зіставність освітніх даних різних країн.
- **Модель стандартизації FAIR-даних для галузі освітніх наук** – розроблена в межах НДР ІЦО НАПН України модель, що охоплює цільовий, змістовий, процесуальний і результативний блоки та містить систему з 15 показників, за якими оцінюють відповідність набору даних принципам FAIR і визначають один із чотирьох рівнів відповідності (див. підрозділ 1.4.3).
- **Набір даних (масив даних; dataset)** – сукупність пов'язаних даних одного дослідження, що зберігається й публікується як окремий ресурс разом з описом (метаданими).
- **Навчальна аналітика (learning analytics)** – збирання й аналіз даних про діяльність здобувачів освіти в цифрових навчальних середовищах (наприклад, журналів активності в системі управління навчанням) з метою розуміння й поліпшення процесу навчання.
- **Онтологія** – формалізований опис понять певної галузі та зв'язків між ними, придатний для автоматичного опрацювання комп'ютером. Онтології дають змогу програмам «розуміти», що різні позначення (наприклад, «клас» і «рік навчання») означають одне й те саме.
- **Оцінювання FAIR-відповідності (FAIRness)** – визначення того, наскільки набір даних відповідає принципам FAIR, за допомогою системи

показників, моделей зрілості або автоматизованих сервісів (наприклад, F-UJI).

- **Пакет FAIR-даних** – завершена сукупність матеріалів дослідження, підготовлена до публікації: дані, метадані, словники, дослідницький інструментарій, дидактичні матеріали, супровідна документація та ліцензія.
- **Пакування даних** – об'єднання файлів даних, їх опису та супровідної документації в один упорядкований пакет для зберігання чи передавання (наприклад, за стандартами RO-Crate або BagIt).
- **Персональні дані** – будь-які відомості, за якими можна прямо або опосередковано встановити особу: прізвище, дата народження, адреса, фото, голос, а також поєднання ознак (заклад освіти, клас, вік, стать), що дає змогу впізнати конкретного учня чи вчителя.
- **План управління даними (Data Management Plan, DMP)** – документ, який дослідник складає ще до початку збирання даних і в якому описує, які дані буде зібрано, як їх буде описано, де і як довго їх зберігатимуть, як захищатимуть персональні дані та на яких умовах дані оприлюднять.
- **Повторна ідентифікація** – встановлення особи учасника дослідження за знеособленими даними шляхом їх зіставлення з іншими відомостями. Ризик повторної ідентифікації особливо високий для малих вибірок і детальних даних.
- **Повторне використання даних** – використання вже зібраних та опублікованих даних для нового дослідження, перевірки результатів, метааналізу чи навчання з дотриманням умов ліцензії та посиланням на джерело.
- **Постійний ідентифікатор (Persistent Identifier, PID)** – незмінне унікальне позначення цифрового об'єкта, особи чи організації, за яким їх можна знайти навіть через багато років: DOI – для набору даних чи публікації, ORCID iD – для дослідника, ROR ID – для установи.
- **Походження даних (provenance)** – відомості про те, звідки взялися дані, як їх зібрано та які зміни (очищення, перекодування, об'єднання) з ними відбулися до публікації.
- **Принципи FAIR** – міжнародні рекомендації щодо організації дослідницьких даних, запропоновані 2016 р. групою науковців на чолі з М. Вілкінсоном. Назву утворено з перших літер англійських слів Findable,

Accessible, Interoperable, Reusable – відшукуваність, доступність, сумісність, багаторазовість використання.

- **Пропрієтарний формат** – формат файлу, що належить певному розробникові програмного забезпечення й повноцінно відкривається лише його програмами (наприклад, формат SAV програми SPSS). Для довготривалого зберігання дані в таких форматах рекомендовано дублювати у відкритих форматах.
- **Псевдонімізація** – заміна прізвищ та інших прямих ідентифікаторів учасників дослідження кодами (наприклад, «Учень 017»), за якої таблицю відповідності кодів і прізвищ зберігають окремо й захищено. Псевдонімізовані дані залишаються персональними.
- **Репозитарій даних** – онлайн-сховище, у якому дослідники зберігають, описують і публікують набори даних і яке забезпечує їм постійний ідентифікатор і довготривалий доступ. Розрізняють загальнонаукові (Zenodo, Figshare, OSF), галузеві та інституційні репозитарії (наприклад, Електронна бібліотека НАПН України).
- **Рівні відповідності FAIR** – чотири рівні (недостатній, базовий, достатній, високий), до одного з яких відносять набір даних за сумою балів за 15 показниками моделі стандартизації FAIR-даних для галузі освітніх наук (див. табл. 1.1).
- **Розпорядник даних (data steward)** – фахівець установи, який консультує дослідників з питань управління даними: допомагає скласти план управління даними, обрати формат, репозитарій і ліцензію, перевіряє якість опису даних.
- **Синтетичні дані** – штучно створені дані, які зберігають статистичні властивості реальних даних, але не містять відомостей про реальних осіб; їх використовують, коли оприлюднення справжніх даних неможливе через ризик ідентифікації учасників.
- **Стандарт метаданих** – узгоджений набір обов'язкових і рекомендованих полів для опису даних (наприклад, Дублінське ядро, DataCite, DDI), що забезпечує однаковий опис даних у різних репозитаріях.
- **Стандартизація даних** – приведення структури, форматів і опису даних у відповідність до загальноприйнятих стандартів, що робить дані зрозумілими й придатними для обміну.

- **Структуровані дані** – дані, упорядковані за чіткою схемою, наприклад таблиця, у якій кожен рядок відповідає респондентові, а кожен стовпець – змінній.
- **Сумісність (Interoperable; інтероперабельність)** – третій принцип FAIR: дані мають бути подані в загальноприйнятих форматах і описані з використанням спільних стандартів і словників, щоб їх можна було поєднувати з іншими даними та опрацьовувати різними програмами.
- **Таксономія FAIR-даних для наукових досліджень у галузі освіти** – розроблена в межах НДР ІЦО НАПН України класифікація освітніх даних за цільовим призначенням дослідження, спеціальністю, суб'єктом подання даних, типами й методами досліджень, типом і рівнем освіти, джерелами, форматом і рівнем агрегованості даних; додатково дані класифікують за режимом доступу, ступенем ідентифікації та етапом життєвого циклу.
- **Файл README** – текстовий файл, що додається до набору даних і коротко пояснює, що містить набір, як організовано файли, як отримано дані, як і на яких умовах їх можна використовувати.
- **Формування FAIR-даних у галузі освітніх наук** – цілеспрямований, методично організований процес збирання, опрацювання, документування, зберігання й (за можливості) публікації освітніх дослідницьких даних протягом життєвого циклу дослідження відповідно до принципів FAIR; важливими складовими є завчасне планування роботи з даними, етичне та правове врегулювання доступу, стандартизований опис і зрозумілі умови повторного використання, можливість інтеграції у подальші наукові пошуки та освітню практику.
- **Цитування даних** – посилання на використаний набір даних у списку джерел статті чи звіту за тими самими правилами, що й на публікацію, обов'язково з постійним ідентифікатором (DOI).
- **Цифровий ідентифікатор об'єкта (Digital Object Identifier, DOI)** – постійний ідентифікатор, який репозитарій присвоює опублікованому набору даних (а видавець – статті); за DOI ресурс завжди можна знайти, навіть якщо змінилася його вебадреса.

Стандарти, сервіси та скорочення латиницею

- **BagIt** – простий стандарт пакування файлів даних для передавання й зберігання: файли розміщують в окремій теці разом зі списком файлів і контрольними сумами, що дає змогу перевірити, чи не пошкоджено дані.
- **DataCite** – міжнародна організація, яка забезпечує присвоєння DOI дослідницьким даним, і однойменний стандарт метаданих, який використовує більшість репозитаріїв даних.
- **DCAT (Data Catalog Vocabulary)** – рекомендований Консорціумом Всесвітньої павутини (W3C) словник для опису наборів даних у каталогах і на порталах відкритих даних.
- **DDI (Data Documentation Initiative)** – стандарт метаданих для докладного опису даних соціологічних, освітніх та психолого-педагогічних опитувань: змінних, питань анкети, вибірки.
- **DMP** – див. План управління даними.
- **DOI** – див. Цифровий ідентифікатор об’єкта.
- **EOSC –** див. Європейська хмара відкритої науки.
- **F-UJI** – безкоштовний онлайн-сервіс, який за DOI автоматично перевіряє, наскільки опублікований набір даних відповідає принципам FAIR.
- **FAIR** – див. Принципи FAIR.
- **GDPR** – див. Загальний регламент про захист даних.
- **ISCED** – див. Міжнародна стандартна класифікація освіти.
- **JSON-LD** – формат запису метаданих, зручний для автоматичного опрацювання й розпізнавання пошуковими системами; його використовують, зокрема, Schema.org і RO-Crate.
- **LOM (Learning Object Metadata)** – стандарт IEEE для опису навчальних ресурсів (цільова аудиторія, рівень складності, тривалість, педагогічне призначення тощо).
- **LRMI (Learning Resource Metadata Initiative)** – доповнення до словника Schema.org для опису навчальних ресурсів так, щоб їх могли знаходити пошукові системи.
- **OAI-PMH** – протокол, за яким агрегатори автоматично збирають метадані з репозитаріїв; його підтримка робить дані репозитарію видимими в загальних пошукових сервісах.

- **ORCID iD** – безкоштовний постійний ідентифікатор науковця (вигляду 0000-0002-1825-0097), що пов’язує з ним усі його публікації та набори даних.
- **PID** – див. Постійний ідентифікатор.
- **RDA FAIR Data Maturity Model** – модель оцінювання FAIR-відповідності, розроблена Альянсом дослідницьких даних (Research Data Alliance, RDA), що містить перелік показників і рівні їх виконання.
- **RDF (Resource Description Framework)** – модель подання даних у вигляді простих тверджень «об’єкт – властивість – значення», що є основою онтологій і дає змогу програмам поєднувати дані з різних джерел.
- **re3data** – міжнародний реєстр репозитаріїв дослідницьких даних, який допомагає дібрати відповідне сховище за галуззю, країною чи умовами доступу.
- **RO-Crate** – сучасний стандарт пакування дослідницьких даних, за яким до набору файлів додають один файл з описом усього пакета (автори, ліцензія, зміст файлів).
- **ROR ID** – постійний ідентифікатор наукової організації в реєстрі Research Organization Registry; дає змогу однозначно зазначати установу автора незалежно від варіантів написання її назви.
- **Schema.org** – спільний словник опису вмісту вебсторінок, створений провідними пошуковими системами; розмітка сторінки набору даних за Schema.org робить його видимим у Google Dataset Search.
- **xAPI (Experience API)** – специфікація запису даних про навчальну діяльність у формі «хто – що зробив – з яким результатом» (наприклад, «учень пройшов тест і отримав 85 %»), яку використовують у системах навчальної аналітики.

Довідкове видання

ФОРМУВАННЯ ТА ВИКОРИСТАННЯ FAIR-ДАНИХ У ГАЛУЗІ ОСВІТНІХ НАУК

Довідник

Керівник прикладного наукового дослідження – **Іванова Світлана Миколаївна,** кандидат педагогічних наук, старший дослідник, завідувач відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України.

Загальна редакція: ***Спірін О. М., Олексюк В. П.***

Авторський колектив:

Спірін Олег Михайлович	доктор педагогічних наук, професор, академік НАПН України, Заслужений діяч науки і техніки України, головний науковий співробітник відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України *(розділ І, п. 1.4; розділ ІІ. п. 2.5; висновки)*
Іванова Світлана Миколаївна	кандидат педагогічних наук, старший дослідник, завідувач відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України *(передмова, розділ І, п. 1.2; розділ ІІ, п. 2.1)*
Вакалюк Тетяна Анатоліївна	доктор педагогічних наук, професор, провідний науковий співробітник відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України *(розділ І, п. 1.7; розділ ІІ, п. 2.4)*
Вознюк Наталя Василівна	молодший науковий співробітник відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України *(розділ І, п. 1.8; розділ ІІ, п. 2.3)*
Кільченко Алла Віленівна	науковий співробітник відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України *(розділ І, п. 1.1; розділ ІІ, п. 2.1; глосарій)*
Мінтій Ірина Сергіївна	доктор педагогічних наук, старший дослідник, доцент, провідний науковий співробітник відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України *(розділ І, п. 1.7; розділ ІІ, п. 2.2; розділ ІІ, п. 2.4)*
Новицька Тетяна Леонідівна	науковий співробітник відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України *(розділ І, п. 1.8; розділ І, п. 1.10; розділ ІІ, п. 2.2)*
Олексюк Василь Петрович	доктор педагогічних наук, професор, старший дослідник, провідний науковий співробітник відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України *(розділ І, п. 1.4; розділ І, п. 1.5; розділ ІІ, п. 2.5)*
Осадча Катерина Петрівна	доктор педагогічних наук, професор, провідний науковий співробітник відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України *(розділ І, п. 1.6; розділ ІІ, п. 2.2; розділ ІІ, п. 2.3)*
Шиненко Микола Андрійович	завідувач сектору відділу відкритих освітньо-наукових інформаційних систем Інституту цифровізації освіти НАПН України *(розділ І, п. 1.9; розділ ІІ, п. 2.3)*

Оформлення обкладинки та комп'ютерна верстка Іванова С. М., Кільченко А. В.

Інститут цифровізації освіти
Національної академії педагогічних наук України
м. Київ, вул. Максима Берлінського, 9
Свідоцтво про державну реєстрацію:
Серія ДК № 7609 від 23.02.2022 р.
електронна пошта (E-mail): iitzn_apn@ukr.net