Олена ЧМИР<sup>1</sup>

## ДОСТУПНІСТЬ ДАНИХ ДОСЛІДЖЕНЬ — «MUST-HAVE» СУЧАСНОЇ НАУКИ

*Розглядаються питання, які мають неперехідну цінність для академічного середовища. Щоб дослідники могли покладатися на наукові результати своїх колег, вони повинні їм довіряти. Тому на передньому плані наукової комунікації стоїть можливість оперативно дізнаватись про успіхи та невдачі інших науковців, отримувати максимально повне уявлення про те, як саме вони були досягнуті, перевірити ці результати, за необхідності повторивши дослідження, коректно посылатися на першоджерела, віддавши данину поваги ученим, роботи яких сприяли науковому прогресу. Що більш доступною буде така інформація, зокрема — не лише короткі повідомлення про наукові досягнення (тези, доповіді чи статті), але й рукописи з детальним висвітленням процесу дослідження, зібраних даних, способів їх обробки та інтерпретації, — то вищою буде ефективність дослідницького процесу. Це принесе зиск не лише науці та освіті, а й суспільству в цілому, яке витрачає на підтримання дослідницького процесу багато коштів і очікує у відповідь продуктивні результати у вигляді нових знань та інновацій. На прикладі наукової спадщини доктора економічних наук, професора, заслуженого діяча науки і техніки Василя Івановича Пили подана історична ретроспектива цінності роботи з даними. Зроблений короткий огляд технічних нововведень (інтернет, загальна комп'ютеризація, електронні наукові та освітні архіви, цифровий формат інформації), які на практиці дозволяють реалізовувати концепцію відкритої науки. Показано, як відбувається імплементація відкритого доступу та*

<sup>1</sup> докторка економічних наук, професорка, завідувачка відділу формування академічних ресурсів, Державна наукова установа «Український інститут науково-технічної експертизи та інформації» (м. Київ), [ES@ukrintei.ua](mailto:ES@ukrintei.ua), <https://orcid.org/0000-0002-9376-7344>.



*відкритої науки в Україні. Викладено авторський погляд на те, що потрібно робити для забезпечення доступності наукових даних максимально широкій аудиторії — представникам академічного середовища (ученим, викладачам), а також усім зацікавленим особам (бізнес, громадянське суспільство). Пропонується запровадити практику складання планів управління даними досліджень та їх опублікування разом із пакетами самих даних і супровідною науковою документацією. Для цього слід задіяти ресурси Національного репозитарію академічних текстів та інституційних репозитаріїв наукових установ і закладів вищої освіти. Правило відкритості даних та планів управління ними має жорстко виконуватись щодо досліджень, на які витрачаються бюджетні кошти. Йдеться про фундаментальні та прикладні наукові дослідження і розробки, а також дисертаційні дослідження здобувачів вищої освіти третього та четвертого рівня — освітньо-наукового та наукового. Подані рекомендації щодо мінімального набору вимог до змісту плану управління даними та процедури їх опублікування, включно з вибором е-архіву, метаописом набору даних, підтриманням доступності й незмінності даних упродовж усього їхнього життєвого циклу.*

**Ключові слова:** наукові дослідження, відкрита наука, репозитарій, е-інфраструктура, дослідницькі дані, управління дослідницькими даними, план управління дослідницькими даними, FAIR.

**В** офіційних списках наукових праць, які містяться, зокрема, у біографічному виданні до 80-річчя від дня народження Василя Івановича Пили [1, с. 27–29] і тематичному номері «Університетських наукових записок» за 2017 рік [2, с. 381–393], присвяченому 85-річчю Василя Івановича, ми знаходимо декілька надзвичайно важливих, але незвичних за форматом публікацій. Це масив фактичних даних щодо наявних лісових ресурсів та розміщення лісової промисловості України, зібраних Василем Івановичем під час написання кандидатської дисертації [3], та інформація з планування розвитку лісопромислового комплексу, підготовлена до 2-го засідання групи експертів ЮНІДО з промислового планування [4].

Фахівці завжди знали ціну первинних матеріалів, особливо тих, що скрупульозно збирались поза основною статистичною звітністю і призначались для вирішення конкретних наукових і практичних завдань. Цими матеріалами ділились з колегами, за ними із відповідними листами-дорученнями їздили до бібліотек наукових установ і відомств. Але широкому загалу ці матеріали, як правило, були недоступні і до бібліографічних переліків та списків наукових праць включались неохоче. Насправді вони представляють неабияку цінність. Доступ до них, — своєчасний та відкритий, — міг позбавити інших дослідників багатьох непродуктивних зусиль, адже вони, замість витратити час на самостійний пошук відповідної інформації, могли б відразу розпочинати свою роботу з аналізу зібраних іншими дослідниками даних. Це була б економія фінансових і трудових ресурсів. Але у період, про який йдеться, мережа інтернет, яку передвіщали Н. Тесла,



Е. М. Форестер, А. Азімов та інші науковці, мрійники і філософи, лише формувалась, а концепція відкритості у нинішньому її розумінні навіть не зародилась. До виникнення та масового поширення всесвітньої мережі<sup>2</sup> залишалось 20 років, до ери доткомів — 25, до інституалізації відкритої науки — 45 років.

Повертаючись до згаданих вище публікацій Василя Івановича [3–4], хочеться зазначити: величезні масиви інформації, яку цілеспрямовано і ретельно він збирав, працюючи над проблемами лісогосподарського, міжгалузевого та регіонального розвитку, так і залишились поза увагою широкого кола фахівців. Водночас навіть коротке ознайомлення з працями Василя Пили з початку 1970-х до кінця 1980-х років [2, с. 381–383] демонструє, з яким цінним масивом фактичних даних, зібраних з різних джерел та старанно упорядкованим, мав справу цей науковець.

Давнє знайомство з ювіляром дозволяє мені розкрити деякі цікаві аспекти його біографії, пов'язані з тематикою цієї статті. Під час свого навчання у Київському лісогосподарському інституті / Українській сільськогосподарській академії у першій половині 1950-х років Василь Іванович мав гарні успіхи у вивченні вищої математики, статистики, лісоводства, економіки, загальної та аналітичної хімії, метеорології, геології, ґрунтознавства, фітопатології, меліорації, лісоустрою та лісоексплуатації. Він жартує у колі друзів, що «отримав спортивну освіту з лісовим нахилом». Адже надзвичайно багато часу приділяв спорту в цілому та баскетболу зокрема. Він створив і очолив інститутську баскетбольну команду і разом з нею досяг стількох перемог, що його намагались перехопити інші університети Києва. Тим не менш, залікова книжка та додаток до диплому інженера лісового господарства № 327991, який він отримав у липні 1955 року, засвідчують, що студент Пила вже тоді цінував точні науки. Пізніше, коли за направленням він почав працювати на великому виробництві в Івдельському районі Свердловської області інженером, старшим інженером, начальником виробничо-технічного відділу та головним інженером підприємства, ці знання стали йому в нагоді. Василю Івановичу довелося займатись ліською таксацією, нормуванням праці, а потім навіть визначати перспективи розвитку крупного виробничого комплексу, обраховуючи наявні запаси лісової сировини та можливості її практичного видобутку на десятки років наперед.

Інтерес до математичного моделювання підсилювався, коли Василь Іванович вступив у 1965 році до аспірантури Ради по вивченню продуктивних сил. Там він познайомився з одним з найвідоміших на той час фахів-

<sup>2</sup> Як відомо, принципи, на яких будується сучасний інтернет, були розроблені та реалізовані на практиці у мережі ARPANET від DARPA (1969), розвинуті у 1980-х роках під час формування NSFNET для Національного фонду досліджень США, задіяні ЦЕРНОм та реалізовані в масовому комерційному сегменті у першій половині 1990-х років. Поява чатів (1988) створила можливість онлайн-спілкування у режимі реального часу, що підвищило популярність інтернету. А запуск веббраузера та розвиток пошукових систем на початку 1990-х років зробили цю мережу незамінною для комунікації мільйонів людей у всьому світі.



ців з лісового господарства та лісової промисловості, який став його науковим керівником — Васильєвим Прокопійем Васильовичем<sup>3</sup>, доктором економічних наук, професором, заслуженим діячем науки, одним із засновників першого в СРСР Інституту лісу.

Написання кандидатської дисертації «Ефективність розвитку та розміщення лісової промисловості Південно-Західного економічного району» [5] у 1965–1969 роках було б неможливе без опрацювання великих масивів даних, які кропітко збирав Василь Іванович. Це первинні дані та упорядкована належним чином статистична інформація щодо запасів лісів, лісоексплуатації, територіальної структури вивезення деревини, сортаментної структури деревини, основних фондів, валової продукції, чисельності працівників на підприємствах лісової та деревообробної промисловості, випуску основних видів продукції лісової промисловості (пиломатеріалів, фанери, ДСП, ДВП, паперу, картону, меблів), а також показники виходу продукції та коефіцієнти переробки деревини, структура споживання лісових матеріалів/деревини за видами виробництв, норми витрат деревини на виробництво різних видів продукції, обсяги попиту в розрізі галузей та видів продукції, економічна доступність лісосировинних ресурсів, транспортні витрати в лісовій промисловості, ефективність виробництва найважливіших видів продукції та ін.

Навички роботи з даними в нашого ювіляра укріплювались, а бажання займатись моделюванням виробничих та економічних процесів — зростало. Це наочно прослідковується в публікаціях Василя Пили. Але унікальні вихідні матеріали дослідження зберігались у його особистому архіві. Пізніше, коли на початку 1980-х років Василь Іванович відчув у собі сили розпочати підготовку докторської дисертації на тему «Програмно-цільове планування міжгалузевого промислового комплексу» [6], трапилась цікава та повчальна історія. Ідея, яка не відпускала науковця весь час та займала його думки — як обрахувати оптимальне розміщення виробничих об'єктів лісогосподарського комплексу з урахуванням максимально повного набору даних — усіх видів ресурсів, транспортних витрат, логістики, виробничих ланцюжків тощо. Василь Іванович поринув у цю ідею та витрачав на її реалізацію увесь свій час, збирав з різних джерел первинні дані підприємств, підбирав та аналізував статистику, знаходив відомчі бази даних, спілкувався з експертами. Декілька відомих фахівців, математиків з Київського національного університету імені Тараса Шевченка, Обчислювального центру Держплану УРСР, провідні науковці Інституту кібернетики ім. В. М. Глушкова НАН України консультували Василя Івановича й пропонували своє бачення постановки задачі з моделювання. Рівень

<sup>3</sup> Васильєв Прокопій Васильович (01.07.1903–14.02.1974). Один із засновників і керівників (1943–1959) Інституту лісу Академії наук СРСР, автор економічного обґрунтування питань захисного лісорозведення (1948–1953), член вченої ради Міністерства лісового господарства РСФСР, завідувач сектору лісового господарства та лісової промисловості Ради по вивченню продуктивних сил при Держплані СРСР (1959–1974), завідувач кафедри Московського лісотехнічного інституту. Науковий керівник та консультант 35 аспірантів і докторантів. Автор понад 200 наукових праць з питань розширеного відтворення лісових ресурсів, системної оцінки лісів та промислового використання деревини.



їхньої експертизи не підлягав сумніву, отже, стоси таблиць з первинними даними, які старанно збирав Василь Іванович, зростали на очах. Поступово ставало очевидно, що всі концепції моделювання, задля яких збирались дані, не спрацьовують. Урешті-решт Василь Іванович не витримав і вирішив порадитись з ще одним фахівцем та своїм гарним знайомим Олександром Олександровичем Бакаєвим<sup>4</sup> [7–8]. Той уважно вислухав історію наукових пошуків, ретельно оглянув зібрані дані, які займали увесь простір великого робочого кабінету Василя Пили, — столи, шафи, підвіконня і зробив невтішний висновок. «Послухай, ти зробив просто титанічну роботу. Я навіть не уявляв, що можна це все віднайти й зібрати! Але, на жаль, сучасні можливості обчислювальної техніки не дозволяють нам звести всі ці дані воедино та урахувати усі обрані тобою чинники виробництва в одній моделі». Та у відповідь на здивування і розпач Василя Івановича, які неможливо було приховати, додав: «Якби це дійсно можна було б змоделювати сьогодні, повір, на цих матеріалах відразу би з'явився як мінімум один академік, декілька членів-кореспондентів, півдесятка докторів наук, а кандидатів — годі й рахувати ...».

Отже, невідповідність між польотом наукової думки і реальністю, що виявилась у рівні доступних технологій<sup>5</sup>, призводила до таких розривів між ідеями і практикою, які на той час неможливо було подолати. Як наслідок, препринт доповіді [4] зберігався в обмеженому колі бібліотек, а дані щодо наявних лісових ресурсів та розміщення лісової промисловості України [3] — у бібліотеках та архівному фонді декількох установ — Держплані УРСР, Міністерстві лісового господарства України та Науково-дослідному економічному інституті. А унікальні бази даних щодо регіонального розвитку та лісопромислового комплексу, зібрані у 1980-х і на початку 1990-х років, були доступні лише вузькому колу фахівців, знайомих з Василем Івановичем особисто.

Але прогрес не стоїть на місці. Сьогодні ми знаємо, що доступ кожного з нас до інформації практично не обмежений завдяки цифровим технологіям та інтернету. Національна або міжнародна статистика? Достатньо скористатись інтернет-ресурсами статистичних відомств окремих держав та міжнародних організацій [9–18], спеціалізованими платформами [19–20], — і ми отримуємо доступ до величезних масивів даних. Звіти про наукові дослідження, рукописи або «сіра література» надзвичайної цінності та актуальності. Доступ до повних текстів 107 945 звітів про НДДКР і 108 409 дисертаційних досліджень надає усім зацікавленим особам Національний репозитарій академічних текстів [21]. Різноманітні дані, зібрані

<sup>4</sup> Бакаєв Олександр Олександрович (11.04.1927–28.11.2009) — український учений-економіст, відомий науковець, праці якого присвячені економіці транспорту, алгоритмізації процесів планування і проектування транспортних систем. Завідувач відділу Інституту кібернетики НАН України (1965), завідувач відділу Міжнародного науково-навчального центру інформаційних технологій та систем НАН та МОН України (1997), академік АН УРСР (1990).

<sup>5</sup> Перші суперкомп'ютери з'явилися у середині 1960-х років, наступний рівень їхньої продуктивності було досягнуто у 1983, чергове зростання потужності відбулось у 1994 та 2008 роках.



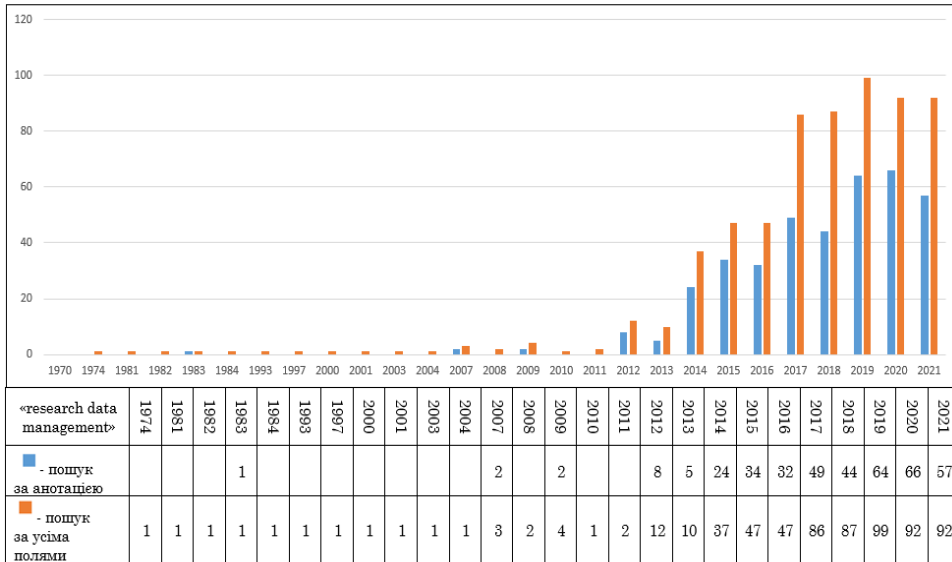
під час проведення наукових досліджень, — медичних, соціологічних, у галузі ядерної фізики, хімії, біології, вивчення ДНК, геопросторові дані. Достатньо перейти на універсальні або вузькоспеціалізовані ресурси відповідного змісту. І можна працювати з матеріалами дослідників з усього світу [22–24].

Про такі можливості годі було й мріяти у 1960–1980-х роках, але наразі це реальність. Ключовими стовпами цієї «нової нормальності» стали інтернет, персональні комп'ютери, концепція відкритості інформації та відомий тепер усім принцип FAIR. Інтернет об'єднав користувачів з усього світу, подолавши географічні, відомчі та часові кордони [25]. Поширення персональних комп'ютерів, які втратили роль унікального робочого приладу і стали звичайним предметом повсякденного життя, долучило спершу сотні тисяч, а нині — мільярди людей до цифрових технологій. Так, за даними Gartner, у 2021 році у світі налічувалось 6,2 млрд персональних комп'ютерів (ноутбуків, настільних ПК, планшетів і мобільних телефонів) [26]. Відкритість даних — адміністративних, публічних, наукових — стала тим «граалем, винайдення якого може призвести до величезного економічного, соціального та політичного прогресу» [27]. З цієї відкритості даних виросла концепція «відкритої науки» як «інклюзивної конструкції, що поєднує різні рухи та практики, спрямовані на те, щоб зробити наукові знання відкритими, доступними і придатними для повторного використання всіма, збільшити наукову співпрацю та обмін інформацією на благо науки і суспільства, відкрити процеси створення, оцінювання та передачі наукових знань за межі традиційної наукової спільноти» [28; 29]. А коли це поєдналось у контексті академічної сфери, закономірно оформилось розуміння ключової цінності відкритої науки — принципів чесності або справедливості, відомих під акронімом FAIR (*Findability, Accessibility, Interoperability, Reusability*). Йдеться про можливість легкого віднайдення, доступності, сумісності, повторного використання та моніторингу результатів досліджень, заснованих на даних [30]. І тепер для кожного науковця стало природньо ділитись з колегами інформацією на всіх етапах дослідження, — від формулювання гіпотези (дискусії у соцмережі Twitter або на платформі Researchgate) до збору, упорядкування та оприлюднення первинних даних, обговорення мотивів обрання того чи того програмного інструментарію для їх обробки, побудови інфографіки, формулювання умовиводів (виявлення закономірностей, підтвердження існування причинно-наслідкових зв'язків) та опублікування на їх основі наукових праць (статей, монографій).

Тут ми безпосередньо підходимо до надзвичайно важливого по суті і не зовсім простого з позицій практичного втілення процесу управління дослідницькими даними<sup>6</sup> (УДД).

Перші публікації з УДД на платформі Web of Science знаходимо на початку 1970-х років [31]. Поступово інтерес до теми зростав, накопичувався досвід, формувалась спільнота прихильників (рис. 1).

<sup>6</sup> Усталена термінологія англійською — *research data management* (RDM).



**Рис. 1. Динаміка публікацій з питань управління дослідницькими даними на платформі Web of Science**

*Примітка:* складено авторкою.

Наразі охочі ознайомитись із літературою щодо управління дослідницькими даними мають обширні можливості: цими питаннями опікується ООН / ЮНЕСКО [32], ОЕСР [33], ЄС [34].

Управління дослідницькими даними є необхідним елементом сучасної науки, інструментом підвищення довіри до неї, ефективним засобом укріплення доказових практик. Наразі у світі державні органи та органи місцевого самоврядування, бізнес-структури, різноманітні фонди спрямовують значні фінансові ресурси на підтримку наукових досліджень. Суттєва частина цих коштів генерується у вигляді податків громадян, відповідно результати досліджень мають ставати якомога більш відкритими, набувати форми суспільного надбання. Видимість результатів досліджень для представників академічної спільноти є важливим чинником, завдяки якому науковці можуть спиратись на раніше отримані іншими дослідниками результати (після здійснення відповідної перевірки / верифікації) й підійматись на більш високу сходинку на старті власного проекту. Адже слова «Якщо я бачив далі інших, то це тому, що стояв на плечах гігантів», відомі нам від Бернара Шартрського, Іоана Солсберійського та Ісаака Ньютона, насправді дуже глибокі та правдиві. Видимість і доступність результатів досліджень для фінансуючих організацій та бізнесу забезпечує їм економію ресурсів й точність постановки завдань: вони можуть у прозорому середовищі ефективно аналізувати увесь дослідницький ландшафт і за результатами такого аналізу обирати найкращі наукові колективи та відш-



ліфувати ідеї для замовлення фундаментальних чи прикладних досліджень. Сказане робить управління дослідницькими даними обов'язковою складовою будь-якого сучасного наукового дослідження. Отже, Україна має імплементувати УДД, наслідуючи європейські підходи та кращі зарубіжні практики.

Перешкоди на цьому шляху, як завжди буває при упровадженні чогось нового, — неминучі. Вони пов'язані з відсутністю досвіду, методичних рекомендацій, відповідної інфраструктури, дефіцитом у закладах вищої освіти і наукових установах високопрофесійних ІТ-фахівців<sup>7</sup>, відсутністю професійних стандартів та освітніх програм для підготовки стюартів даних. Водночас відкритість наукових даних має бути забезпечена. Це — «must-have» сучасної науки, тому новий виклик академічна спільнота має прийняти і достойно на нього відповісти.

Що для цього потрібно? Частину зазначених питань намагається вирішити робоча група з розроблення національного плану щодо відкритої науки, створена Міністерством освіти і науки України [35]. Членами цієї групи у другій половині 2021 року розроблений проект Національного плану відкритої науки, який передбачає розвиток відкритого доступу до наукових публікацій, упровадження принципів FAIR, розбудову дослідницьких інфраструктур, підтримку громадянської науки, реформування освітньої компоненти тощо. У частині просування відкритого доступу пропонується забезпечити відповідне поширення інформації в академічній спільноті, консультування відповідальних осіб закладів вищої освіти і наукових установ щодо розвитку інституційних репозитаріїв, підтримку видання журналів відкритого доступу, запровадження відкритого рецензування та механізмів моніторингу прогресу у цій сфері. Щодо відкритих та FAIR даних, відкритого програмного забезпечення й підтримки відтворюваних досліджень, які вбачаються членами робочої групи надзвичайно важливим інструментом ефективної науки, то ставляться завдання щодо створення нормативно-організаційної бази розвитку інфраструктури й компетенцій для роботи з дослідницькими даними. Для цього визнається за необхідне реалізувати цілий комплекс заходів, серед яких — запровадження вимог щодо FAIR даних; зобов'язання щодо обов'язкового складання та виконання під час НДДКР плану управління дослідницькими даними; розміщення дослідницьких даних у репозитаріях під ліцензією CC BY; визнання дослідницьких даних як наукового результату НДДКР та заохочення їх оприлюднення, повторного використання й цитування; створення організаційних та фінансових передумов для розвитку е-архівів даних, їх належної сертифікації.

Слід зазначити, що поштовх до такого глибокого розгляду питань відкритої науки дала ініціативна група<sup>8</sup>, яка готувала пропозиції до проекту

<sup>7</sup> Внаслідок неконкурентоспроможності цих установ на цьому специфічному ринку праці, де найбажанішими роботодавцями є великі міжнародні високотехнологічні компанії.

<sup>8</sup> У складі: Оксана Бруй, кандидат наук із соціальних комунікацій, президент ГО «Українська бібліотечна асоціація», директор науково-технічної бібліотеки Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»; Олена Чмир, д-р екон. наук, професор, завідувач відділу супроводження академічних ресурсів





плану дій із впровадження Ініціативи «Партнерство «Відкритий Уряд» у 2021–2022 роках. Під час обговорення нагальних проблем академічної сфери було наголошено на необхідності всебічної підтримки відкритої науки [36]. Пропонувалось розробити проєкт Національного плану з відкритої науки, провести його широке публічне обговорення науковцями та керівниками академічних інституцій, узгодити із заінтересованими центральними органами виконавчої влади, підготувати пропозиції щодо внесення змін до діючих нормативно-правових актів<sup>9</sup> [37].

Зважаючи на все вищевикладене і спираючись на матеріали опрацьованих нами численних публікацій з питань відкритої науки та відкритих даних, пропонуємо власне бачення того, як має відбуватись упровадження управління дослідницькими даними (ДД) в Україні.

Під час проведення досліджень, у рамках яких передбачається здійснювати збір, упорядкування та аналіз даних, науковці повинні заздалегідь планувати увесь пов'язаний з цим комплекс робіт (рис. 2).

<ul style="list-style-type: none"> <li>формулювання ідеї наукового дослідження,</li> <li>обговорення у науковому колективі суттєвих рис НДДКР,</li> <li>написання проєктної заявки або технічного завдання на проведення дослідження</li> </ul>	<ul style="list-style-type: none"> <li>обговорення із замовником (грантодавцем, фінансуючою організацією) проєктної заявки або технічного завдання на дослідження,</li> <li>підписання угоди про виконання дослідження</li> </ul>	<ul style="list-style-type: none"> <li>збір даних,</li> <li>упорядкування даних,</li> <li>забезпечення надійного збереження даних,</li> <li>опис даних,</li> <li>супроводження даних необхідною науковою документацією для розуміння дослідницьких даних сторонніми науковцями,</li> <li>забезпечення незмінності даних,</li> <li>резервне копіювання даних</li> </ul>	<ul style="list-style-type: none"> <li>оприлюднення (опублікування, депонування) даних та їхніх метаописів на цифрових платформах чи електронних архівах</li> </ul>	<ul style="list-style-type: none"> <li>забезпечення незмінності даних,</li> <li>підтримання доступності даних,</li> <li>резервне копіювання даних,</li> <li>поширення,</li> <li>оновлення форматів даних та їхніх метаописів</li> </ul>
підготовчі роботи		проведення дослідження	завершення оформлення виконаного дослідження, звітування	збереження та підтримання доступності даних поза межами дослідницького проєкту
складання плану УДД		актуалізація (оновлення) плану УДД		
		виконання плану УДД		

Рис. 2. Загальна схема управління дослідницькими даними

*Примітка:* складено авторкою.

УкрІНТЕІ; Ірина Кучма, експерт проєкту OpenAIRE, координатор програми «Відкритий доступ» у eIFL.net, тренінгової програми проєкту «FOSTER» з розвитку навчальних і освітніх програм з відкритої науки, член Research Data Alliance; Михайло Серебряков, експерт урядово-громадської ініціативи «Разом проти корупції».

<sup>9</sup> Зокрема Закону України «Про науково-технічну інформацію», Постанови Кабінету Міністрів України від 21 жовтня 2015 р. № 835 «Про затвердження Положення про набори даних, які підлягають оприлюдненню у формі відкритих даних», Наказу Міністерства освіти і науки України від 04 липня 2018 р. № 707 «Про затвердження Регламенту роботи Національного репозитарію академічних текстів».



Він має розпочинатись з обговорення членами майбутнього творчого колективу (ще на етапі підготовки грантової заявки) питань збору даних, а також можливості чи необхідності їх оприлюднення та надалі узгоджуватись під час консультацій із замовником (грантодавцем, фінансуючою організацією). Для цього формується і постійно оновлюється така важлива складова дослідницької документації, як план управління дослідницькими даними (ПУДД).

План управління дослідницькими даними є документом, який максимально чітко й однозначно описує весь процес виконання дослідження з позицій поводження з даними: їх збирання, накопичення, обробку, опис, зберігання, оприлюднення. ПУДД визначає загальні принципи, норми, правила й послідовність дій усіх учасників наукового колективу для дотримання високої якості, коректності, точності, валідності, безпеки, незмінності отриманих під час дослідження даних, їх оформлення, опису і поширення таким чином, щоб вони були зрозумілі, доступні й придатні для повторного використання. Такий план, підготовлений і погоджений із замовником, може уточнюватись в міру виконання НДДКР та набуває остаточного вигляду на завершальній стадії дослідження.

Ми вважаємо, що для відкритих досліджень<sup>10</sup>, які виконуються із залученням коштів державного чи місцевих бюджетів України, ПУДД також має бути відкритим. Для цього він оприлюднюється разом із описом очікуваних суттєвих результатів дослідження під час державної реєстрації НДДКР на початку дослідження, а його остаточна версія стає доступною широкому загалу разом із описом отриманих фактичних результатів дослідження<sup>11</sup> [38]. Те саме, але в більш гнучкому форматі мають виконувати здобувачі вищої освіти третього і четвертого рівнів під час своєї роботи над дисертаційним дослідженням. Відмінність буде полягати в тому, що публікація ПУДД разом із пакетом дослідницьких даних та іншою науковою документацією здійснюватиметься не пізніше визначеного терміну оприлюднення дисертації перед її публічним захистом або у процесі оформлення облікової картки дисертації [38].

Бажано, щоб наукова громадськість мала змогу ознайомитись із планами управління даними будь-яких відкритих досліджень, навіть тих, що не були профінансовані із залученням бюджетних коштів. Тому рекомендується всі плани УДД, оприлюднювати, використовуючи для цього інституційні репозитарії закладів вищої освіти і наукових установ [39] та / або Національний репозитарій академічних текстів [21].

У ПУДД надаються загальні відомості (резюме) даних дослідження, інформація про доступність метаописів і власне самих пакетів дослідниць-

<sup>10</sup> Тобто таких, які не мають обмежень щодо доступу до їхніх результатів згідно з чинним законодавством України (не є секретними, не призначені лише для службового використання тощо).

<sup>11</sup> У процесі оформлення реєстраційної картки науково-дослідних, дослідно-конструкторських робіт, облікової картки науково-дослідних, дослідно-конструкторських робіт, які оформлюються згідно з актуальним Порядком державної реєстрації та обліку науково-дослідних, дослідно-конструкторських робіт і дисертацій.



ких даних, пояснення щодо їх повноти та специфіки, місце збереження. Зокрема, у плані вказується, чи будуть дані дослідження повністю відкритими, частково відкритими або закритими, якою буде структура даних, чи це оригінальні, самостійно зібрані під час виконання НДДКР або складені з різних зовнішніх джерел та опрацьовані дані, а також пояснюються підходи до їх упорядкування. Наводяться відомості щодо використання універсальних цифрових ідентифікаторів і стандартів опису метаданих, надається реферативний опис, авторський набір ключових слів. Робляться пояснення щодо тих даних, які не будуть відкритими: чи можуть вони бути надані на індивідуальний запит, чи передбачене відшкодування/оплата за передавання даних зацікавленим особам тощо. Вказується термін ембарго (строк відтермінування оприлюднення), якщо це передбачено у ПУДД, особливості ліцензування закритої частини даних, відомості про правові обмеження згідно з чинним законодавством, інформація про патентний захист, захист персональних даних і засоби його забезпечення. Обов'язково наводиться інформація щодо зобов'язань і тривалості підтримання доступності даних й оновлення форматів даних як у часових рамках виконання дослідження, так і після його завершення. Подається інформація щодо словників, онтологій, методик, робочих зошитів, процедур та описів експериментальних досліджень, використання яких разом із пакетом даних забезпечить його максимально ефективно залучення в науковий обіг. Розкриваються відомості щодо зав'язків з іншими дослідженнями, раніше опублікованими даними, звітами про поточне та попередні дослідження, науковими публікаціями авторів дослідження, де описані результати, отримані на основі роботи з цими даними. Надаються пояснення щодо програмних продуктів, які використовуються для роботи з даними (відкриті чи комерційні/ліцензовані програмні продукти), вказуються альтернативні можливості щодо використання програмного забезпечення з відкритим кодом. Наводяться посилання на державні та міжнародні стандарти, класифікатори, методики, схеми дослідження і збору даних, методи, протоколи збору, упорядкування та опрацьовання даних, використані інструменти та обладнання, подається опис робочих процесів та алгоритмів. У разі необхідності надаються пояснення щодо фінансових, людських, матеріальних ресурсів, які мають бути задіяні на всіх етапах роботи з даними — від підготовки заявки на грант або технічного завдання на проект, до його завершення та подальшого підтримання доступності дослідницьких даних. Сказане стосується як очікуваних (чітко прописаних у технічному завданні чи грантовій угоді) результатів НДДКР, так і тих, що можуть бути отримані під час дослідження неочікувано, спонтанно.

У практичній площині управління дослідницькими даними здійснюється під час проведення робіт зі збору, накопичення, документування, упорядкування, закриття або анонімізації (у разі необхідності), зберігання та зрозумілого для сторонніх осіб представлення даних дослідження на основі принципів FAIR, у процесі їх оприлюднення, опублікування, депонування на обраних цифрових платформах або сховищах, підтримання даних упро-



довж усього життєвого циклу, забезпечення незмінності, резервного копіювання, збереження, поширення, оновлення форматів даних і метаданих навіть поза часовими рамками виконання дослідницького проекту.

Як уже зазначалось, роботи з управління даними розпочинаються на етапі планування дослідження і завершуються після виконання всіх визначених у ПУДД дій, включно з підтриманням доступності даних та оновленням їхніх форматів після завершення дослідження.

Розміщення описів (метаданих) та безпосередньо повних пакетів дослідницьких даних (у комплекті з відповідною науковою документацією — інструктивними матеріалами, методиками, поясненнями, словниками, онтологіями тощо) у наукових сховищах може відбуватись під час проведення дослідження в міру збору та упорядкування даних, але не може затримуватись понад час, обумовлений у грантовій угоді, договорі, технічному завданні чи ПУДД. Практика показує, що публікація ДД під час дослідження робить його більш продуктивним, оскільки виконавці НДДКР можуть отримати зовнішню експертизу від усіх заінтересованих осіб: публічна дискусія, обговорення, рецензування, обмін ідеями з академічною спільнотою та суб'єктами, зацікавленими в результатах дослідження. Публікація ДД на етапі передавання рукопису на розгляд редакційної колегії наукового видавництва перед відповідним рецензуванням робить експертизу більш ефективною. Публікація ДД після повного завершення дослідження фіксує зібрані та упорядковані дані й дозволяє продемонструвати обґрунтованість зроблених висновків і підготовлених пропозицій. В окремих випадках на кшталт надзвичайних непередбачуваних обставин, розкриття ДД може бути прискорене (як це було для медичних даних під час пандемії COVID-19) або загальмоване (у разі природних катастроф, військових конфліктів, руйнування об'єктів наукової інфраструктури, запровадження більш жорстких вимог національної безпеки тощо).

Механізмом оприлюднення та підтримки доступності ДД є депонування на наукових та освітніх цифрових платформах. Це відбувається шляхом самостійного внесення або передавання до уповноваженої інституції.

Базовою інфраструктурою для зберігання дослідницьких даних в Україні ми вбачаємо Національний репозитарій академічних текстів та інституційні репозитарії наукових установ і закладів вищої освіти. Йдеться про дані, отримані під час наукових досліджень, що виконуються в Україні або за участі вітчизняних науковців. Водночас, за погодженням із замовником (грантодавцем, фінансуючою організацією), ДД також можуть депонуватись в інших наукових цифрових сховищах — репозитаріях, які забезпечують надійний і довгостроковий доступ до цифрових ресурсів. Це можуть бути репозитарії мережі EOSC [40], Zenodo [23], рекомендовані дослідницькими спільнотами у рамках певної галузі знань, сховища, що відповідають ISO16363, мають сертифікати CoreTrustSeal, DIN31644, включені до реєстру на порталі re3data.org [41] тощо.

Метаопис дослідницьких даних повинен відповідати актуальним універсальним та галузевим стандартам, а також вимогам сховища, де



вони будуть зберігатись. Як мінімум, він має містити відомості про усіх авторів дослідження, назву пакету даних, назву дослідження, у рамках якого отримано ці дослідницькі дані (з шифром проекту, гранту або номером державної реєстрації НДДКРiДР), анотацію та ключові слова, точну дату початку та завершення роботи з даними, ліцензію, термін ембарго, міжнародні ідентифікатори, зокрема — установ, де виконано дослідження (ROR), авторів (ORCID), самого пакету дослідницьких даних (DOI).

Звісно, описані вище підходи мають рамковий характер. У різних галузях знань, дисциплінах і спеціальностях склались специфічні практики, норми і правила, що найбільш яскраво виявляється у медицині, біохімії, астрономії, фізиці тощо. Ці особливості також потрібно враховувати під час роботи з дослідницькими даними. У будь-якому випадку автори / упорядники несуть відповідальність за цілісність вмісту даних, зв'язок із матеріалами дослідження, зокрема — рукописом, у якому розкривається його хід та основні результати, а також публікаціями, завдяки яким ці результати були презентовані академічній спільноті.

Наразі ми в Україні знаходимось на початку шляху до відкритих даних досліджень. Але багато підготовчих заходів вже реалізовано, є розуміння необхідності такої роботи, політична воля і підтримка науковців та освітян. Будемо сподіватись на швидкий прогрес у цій сфері.

#### Список використаних джерел

1. Василь Іванович Пила. До 80-річчя від дня народження та 45-річчя науково-педагогічної та громадської діяльності: довідково-біографічне видання / упор. Н. П. Захаркевич, Л. Л. Місінкевич. Хмельницький : ВИС, 2012. 70 с. URL : <https://t.ly/ZebrE>.
2. Список публікацій Василя Івановича Пили. *Університетські наукові записки*. 2017. № 1. С. 381–392. *Національна бібліотека України ім. В. І. Вернадського НАН України*. URL : [https://t.ly/AW\\_Dk](https://t.ly/AW_Dk).
3. Пила В. И., Васильев П. В. Справочные материалы по лесным ресурсам и размещению лесной промышленности Украинской ССР. Москва : СОПС при Госплане СССР, 1973. 168 с.
4. Pila V. I. Planning the development of timber industry system / UNIDO-United Nations Industrial Development Organization. Informational report to 2nd meeting of UNIDO expert group on industrial planning. May 21–25, 1984, Kiev.
5. Пила В. И. Эффективность развития и размещения лесной промышленности юго-западного экономического района. Дисс. ... канд. экон. наук / Совет по изучению производительных сил, Государственный плановый комитет СССР. Москва, 1969. 268 л.
6. Пила В. И. Программно-целевое планирование межотраслевого промышленного комплекса. Дисс. ... д-ра экон. наук / Научно-исследовательский экономический институт Госплана Украинской ССР. Киев. 1987. 350 л.
7. Бакаев Александр Александрович (11.04.1927–28.11.2009). *Національна академія наук України*. URL : [https://t.ly/Irq\\_M](https://t.ly/Irq_M).
8. Бакаев Александр Александрович (1927–2009). *Національна бібліотека України ім. В. І. Вернадського НАН України*. URL : [https://t.ly/1M7T\\_](https://t.ly/1M7T_).



9. Державна служба статистики України. URL : <https://www.ukrstat.gov.ua>.
10. The United Nations Statistical Commission. URL : <https://t.ly/YT74E>.
11. Department of Economic and Social Affairs Statistics. Publications. *The United Nations*. URL : <https://t.ly/Vg85M>.
12. UN Big Data. Big Data and Data Science for Official Statistics. *The United Nations*. URL : <https://t.ly/4sB5N>.
13. Find, compare and share the latest OECD data: charts, maps, tables and related publications. *OECD*. URL : <https://www.oecd.org>.
14. IMF Publications. *IMF*. URL : <https://www.imf.org/en/Publications>.
15. World Bank Open Data. Free and open access to global development data. *World Bank Open Data*. URL : <https://data.worldbank.org>.
16. World Bank Research and Publications. *World Bank Open Data*. URL : <https://www.worldbank.org/en/research>.
17. Welcome to Eurostat. *Eurostat*. URL : <https://ec.europa.eu/eurostat>.
18. Research and innovation. *The European Commission*. URL : [https://ec.europa.eu/info/research-and-innovation\\_en](https://ec.europa.eu/info/research-and-innovation_en).
19. The company behind the success. *Statista*. URL : <https://www.statista.com>.
20. Портал відкритих даних. Єдиний державний веб-портал відкритих даних. URL : <https://data.gov.ua>.
21. Національний репозитарій академічних текстів. *HPAT*. URL : <https://nrat.ukrintei.ua>.
22. Resources. CERN Accelerating science. *CERN*. URL : <https://t.ly/-ixMm>.
23. Zenodo. URL : <https://zenodo.org>.
24. Figshare. Repository. *Figshare*. URL : <https://figshare.com>.
25. A short history of the Internet. *National Science and Media Museum, in the heart of Bradford*. URL : <https://t.ly/vwGjG>.
26. Gartner Forecasts Global Devices Installed Base to Reach 6.2 Billion Units in 2021. URL : <https://t.ly/q-jNf>.
27. Doherty D. Jisc futures: does business struggle with the concept of open access? (17.07.2017). *Times Higher Education*. URL : <https://t.ly/UMuIs>.
28. Draft text of the UNESCO Recommendation on Open Science / Intergovernmental Meeting of Experts (Category II) Related to a Draft UNESCO Recommendation on Open Science, online, 2021 SC-PCB-SPP/2021/OS-IGM/WD3. URL : <https://t.ly/Gdktn>.
29. Чмир О. С. Розвиток в Україні ресурсів, заснованих на знаннях. *Наука, технології, інновації*. 2021. № 4. С. 35–49. <http://doi.org/10.35668/2520-6524-2021-4-04>.
30. Wilkinson M., Dumontier M., Aalbersberg I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016. Vol. 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.
31. Baker L. R. An Adaptable Interactive System for Medical and Research Data Management. *Methods of Information in Medicine*. 1974. Vol. 13, Iss. 04. P. 209–215. <https://doi.org/10.1055/s-0038-1636153>.
32. Unesdoc. Literature for «research data management». *UNESCO digital library*. URL : <https://t.ly/vS44b>.
33. Title, Authors or ISSN/ISBN/DOI «research data» AND (Authors «OECD»). *OECD i-library*. URL : <https://t.ly/uHvli>.
34. Practical guide to the international alignment of research data management / Extended Edition with DMP Evaluation Rubric. 2021. *Science Europe*. URL : <https://t.ly/IzNBi>.



35. Про утворення робочої групи з розроблення національного плану щодо відкритої науки. Наказ Міністерства освіти і науки України від 02.06.2021 р. № 614. *Міністерство освіти і науки України*. URL : <https://t.ly/lIpFK>.
36. Запрошуємо до обговорення проекту розпорядження Кабінету Міністрів України «Про затвердження плану дій із впровадження Ініціативи «Партнерство «Відкритий Уряд» у 2021–2022 роках (27.11.2020; 12:25). *Урядовий портал*. URL : <https://t.ly/tRom6>.
37. Про затвердження плану дій із впровадження Ініціативи «Партнерство «Відкритий Уряд» у 2021–2022 роках. Проект розпорядження Кабінету Міністрів України. *Урядовий портал*. URL : <https://t.ly/fcQVc>.
38. Порядок державної реєстрації та обліку науково-дослідних, дослідно-конструкторських робіт і дисертацій. Наказ Міністерства освіти і науки України від 27.10.2008 р. № 977 зі змінами від 09.06.2009 р. № 494. URL : <https://t.ly/YYqS7>.
39. Корисні посилання. *HPAT*. URL : <https://t.ly/UMqVi>.
40. European Open Science Cloud — EU Node URL : <https://t.ly/Sa2wE>.
41. Re3data. Global registry of research data repositories. URL : <https://www.re3data.org>.

Надійшла до редакції 13.01.2022

**Olena CHMYR**

*(Ukrainian Institute of Scientific and Technical Expertise and Information)*

### **Availability of Research Data as a «Must-Have» of Modern Science**

*The article deals with the issues of enduring value to the academic community. Researchers must trust their colleagues to rely on their scientific results. Therefore, the cornerstone of scientific communication is the ability to know the content of colleagues' work (specifically their successes and failures), to understand how exactly these results were achieved, and to verify them by replicating the research, if necessary. Naturally, it's also essential to cite correctly primary sources, giving due respect to the contributions of other scientists, whose work has aided one's progress in science. This means that the higher the availability of manuscripts or other comprehensive descriptions of scientific results with detailed descriptions of the research process, collected data, methods of processing, and interpretation, the higher the efficiency of the research process. This is beneficial not only for science and education but also for society as a whole, which invests a lot of resources in supporting the research process, and expects productive results in the form of new knowledge and innovations. The article provides a retrospective analysis of the value of working with data by using the scientific heritage of Doctor of Economic Sciences, Professor, and Honoured Worker of Science and Technology Vasily Ivanovich Pyla as an example. A brief overview of technical innovations (the internet, universal computerization, electronic scientific and educational archives, and digital information formats) is given in the paper. This translates the concept of open science into practical terms. It shows how open access and open science are being implemented in Ukraine. An author's perspective is offered on how to ensure that the data from scientific research by domestic scientists becomes available to the widest possible audience: representatives of the academic community (scientists, teachers), as well as all other interested parties (business, civil society). It is also proposed to draw up a research data management plan and publish it along with the data packages and accompanying scientific documentation for convenient utilization. In this case, the capabilities of the National Repository of Academic Texts and the institutional repositories of scientific organizations and higher*



---

*education institutions should be used. This approach should be strictly applied for research projects, conducted for budget funds (scientific research and developments, as well as dissertations (thesis) by the applicants of the third and fourth levels of educational-scientific qualifications (degrees). Recommendations are given regarding the content of the data management plan (minimum set of requirements) and the publication procedure itself (including the choice of e-archive, metadata description of the data set, and maintaining data protection availability from hostile changes throughout their lifecycle).*

**Keywords:** *research, open science, repository, e-infrastructure, research data, research data management, data management plan, FAIR.*