

<https://doi.org/10.62405/osiconf.2026.23>

Петренко Оксана,

завідувач науково-організаційного відділу, к.і.н.
Державна науково-технічна бібліотека України, Київ, Україна,
ORCID: 0000-0001-5788-8192, e-mail: lo.petrenko@dntb.gov.ua

Хименко Богдан,

науковий співробітник науково-організаційного відділу, аспірант
Державна науково-технічна бібліотека України, Київ, Україна,
ORCID: 0009-0007-2612-8909, e-mail: bogkhym@gmail.com

РОЛЬ КОНТРОЛЬОВАНИХ СЛОВНИКІВ ДЛЯ РЕПОЗИТАРІЇВ ДОСЛІДНИЦЬКИХ ДАНИХ

АНОТАЦІЯ. Мета роботи — обґрунтувати роль контрольованих словників як базового елемента семантичної архітектури репозитаріїв дослідницьких даних і визначити склад завдань зі створення такої системи для національного репозитарію. Використано документальний та порівняльний аналіз міжнародних стандартів і галузевих словників, контент-аналіз рекомендацій міжнародних організацій, термінологічний аналіз та функціональне моделювання. Установлено шість основних функцій, які контрольовані словники виконують у репозитаріях: забезпечення якості й узгодженості описів, покращення пошуку та виявлення даних, підтримання інтероперабельності, структурування метаданих, довготривале збереження та автоматизація оброблення даних. Систематизовано нормативну базу і схарактеризовано провідні галузеві словники, придатні для інтеграції в репозитарії. Запропоновано склад семантичної інфраструктури національного репозитарію, що охоплює архітектуру словників, формати обміну, програмний інтерфейс, вебінтерфейс і методичне забезпечення. Зроблено висновок, що підтримання словників має бути постійною інституційною функцією, а не разовим проєктом.

КЛЮЧОВІ СЛОВА: контрольовані словники; репозитарій дослідницьких даних; метадані; інтероперабельність; принципи FAIR; відкрита наука.

ВСТУП. Розбудова національних інфраструктур відкритої науки перевела питання семантичної узгодженості наукових даних із вузькоспеціальної бібліотекознавчої площини у площину державної науково-технічної політики. Закон України від 25 лютого 2026 р. № 4794-IX «Про внесення змін до Закону України “Про наукову і науково-технічну діяльність” щодо питань дослідницької інфраструктури та підтримки молодих вчених» уперше на законодавчому рівні визначив поняття «відкрита наука», «відкритий доступ», «дослідницька інфраструктура», «Національна електронна науково-інформаційна система». Зокрема визначення «принципи FAIR — принципи належного управління дослідницькими даними, що передбачають забезпечення багаторазового використання дослідницьких даних, їх доступність, здатність до взаємодії з різними типами даних (інтероперабельність) та здійснення оперативного пошуку необхідної інформації» та «оптимізовані дослідницькі дані (FAIR-дані) — дослідницькі дані, що зберігаються в електронній формі та відповідають принципам Findable, Accessible, Interoperable, Re-usable» засвідчують, що ці поняття є безпосередньо релевантними для усіх науково-технічних досліджень.

Розпорядження Кабінету Міністрів України від 8 жовтня 2022 р. № 892 «Про затвердження Національного плану щодо відкритої науки» та від 17 червня 2026 р. № 603-р «Про затвердження плану дій із впровадження Ініціативи «Партнерство “Відкритий Уряд” у 2026–2027 роках», а також Дорожня карта інтеграції України до Європейського дослідницького простору на період до 2027 року визначили створення Національного репозитарію дослідницьких даних (далі — НРДД) одним із пріоритетних завдань. Проєкт створення та розбудови НРДД виконує ДНТБ України. Наразі розгор-

нуто бета-версію НРДД на платформі Dataverse¹ та підготовлено комплект нормативно-методичних документів. Наступним етапом є пілотне тестування репозитарію, зокрема перевірка стабільності депонування наборів даних і генерації постійних ідентифікаторів тощо. Однак технічна спроможність зберігати набори даних сама собою не забезпечує їхньої наукової цінності. Репозитарій, у якому той самий об'єкт описано десятком неузгоджених термінів, залишається сховищем файлів, а не інструментом наукової комунікації: дані в ньому фізично наявні, але практично невидимі для потенційного користувача.

Для України ця проблема має додатковий вимір. Україномовна наукова термінологія залишається невпорядкованою та недостатньо гармонізованою з міжнародною, що ускладнює уніфікацію наукової інформації і створює бар'єри для її подальшого використання. Наслідки цього цілком практичні: неточний та трудомісткий пошук, ускладнене структурування даних, неповнота метаданих і, як підсумок, обмежена можливість відтворення й повторного використання результатів досліджень. Дослідницькі дані, створені різними установами в різних галузях, часто залишаються фрагментованими саме через розбіжності в термінології, а не через технічні обмеження платформ.

Емпіричне обґрунтування цінності контрольованих словників має вже кількадесятилітню історію. Дослідження T. Gross, A. G. Taylor і D. N. Joudrey (Gross et al., 2015) показало, що від чверті до третини записів, які повертає повнотекстовий пошук, було б утрачено за відсутності контрольованої лексики; принципово, що цей результат зберігся й після збагачення каталожних записів автоматично згенерованими даними на кшталт змісту та анотацій. C. Smith (Smith, 2021) простежує еволюцію предметного доступу і фіксує зміщення від універсальних словників, які підтримують національні бібліотеки та великі професійні організації, до спеціалізованих словників вузького призначення, що супроводжується як розширенням можливостей, так і зростанням ризиків фрагментації.

Стосовно середовища наукових даних показовим є опитування учасників мережі DataNet, проведене Y. Zhang, A. Ogletree, J. Greenberg і C. Rowell (Zhang et al., 2015): дослідники, куратори та розробники репозитаріїв розглядають контрольовані словники як цінний інструмент і водночас потребують репозитарних технологій, які надавали б доступ одразу до кількох словників. A. Roy і S. Ghosh (Roy & Ghosh, 2025) звертають увагу на суттєву прогалину в чинних рекомендаціях: Конфедерація репозитаріїв відкритого доступу (COAR) пропонує репозитаріям три типи контрольованих словників — права доступу, тип ресурсу і тип версії, — але не пропонує предметного словника, хоча саме предметна рубрика належить до найуживаніших пошукових категорій. Аналіз репозитаріїв соціогуманітарного профілю з реєстру OpenDOAR виявив, що предметне впорядкування суттєво різниться від репозитарію до репозитарію, а отже потребує стандартизації задля інтеоперабельності.

Окремий напрям сучасних досліджень стосується автоматизації. C. Foster та колеги (Foster, 2024) застосували гармонізований перехресний показник контрольованих словників (UMLS, термінологія DevTox Федерального інституту оцінювання ризиків Німеччини та кінцеві точки ОЕСР) приблизно до 34 000 і 6 400 витягів даних із досліджень токсикології розвитку та автоматично стандартизували відповідно 75 % і 57 % описів кінцевих точок — результат, який демонструє і потенціал, і межі автоматизованого підходу. J. Saldanha Bach Estevao і C.-P. Klas (Saldanha Bach Estevao & Klas, 2024) описують контрольований словник GESIS для змінних у соціогуманітарних дослідницьких даних, опублікований через менеджер словників CESSDA, що дає змогу експлікувати відношення між змінними різних хвиль опитувань. Водночас L. Vogt зі співавторами (Vogt, 2025) переконливо показують, що самим принципам FAIR бракує конкретності щодо семантичної інтеоперабельності, і пропонують їх розширення на основі лінгвістичної перспективи, а G. Miller і E. Spiegel (Miller & Spiegel, 2025) вписують узгодженість опису змінних у ширшу рамку цілісності дослідницьких даних.

¹<https://uad.dntb.gov.ua/>

Невирішена раніше частина загальної проблеми. Попри розвиненість міжнародної дискусії, поза її межами лишається питання побудови багатомовної семантичної інфраструктури національного репозитарію для мови, наукова термінологія якої систематично не гармонізована з міжнародною. Наявні праці описують або практику окремих англomовних інфраструктур, або галузеві рішення, але не пропонують цілісної моделі, придатної для національного рівня в умовах термінологічної невпорядкованості.

Мета роботи — обґрунтувати роль контрольованих словників як базового елемента семантичної архітектури репозитаріїв наукових даних і визначити склад та послідовність завдань із розроблення й упровадження системи контрольованих словників для Національного репозитарію наукових даних.

МЕТОДИ. Дослідження має аналітично-оглядовий характер і виконане на перетині термінознавства, наукознавства та інформаційних технологій. Використано документальний аналіз чинної нормативно-правової бази України у сфері відкритих даних і відкритої науки, а також корпусу міжнародних стандартів ISO, що регламентують побудову тезаурусів, управління термінологічними ресурсами та структуру метаданих. Порівняльний аналіз застосовано до провідних міжнародних галузевих словників за критеріями предметного охоплення, обсягу термінологічного масиву, підтримуваних форматів подання та умов ліцензування. Контент-аналіз рекомендацій COAR, реєстру BARTOC і матеріалів Research Data Alliance дав змогу виокремити усталені практики імплементації словників у репозитарному середовищі.

Термінологічний аналіз використано для розмежування форм організації контрольованої лексики та визначення співвідношення між ними. Метод функціонального моделювання застосовано для опису складу семантичної інфраструктури репозитарію та зв'язків між її компонентами. Узагальнення й систематизація результатів дали змогу сформулювати перелік функцій контрольованих словників та склад завдань упровадження. Емпіричною базою слугували публікації 2015–2025 рр., індексовані у Scopus і Web of Science, тексти міжнародних стандартів, рекомендаційні документи ЮНЕСКО та Європейської Комісії, а також відкриті описи чинних галузевих словників.

РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ. Контрольований словник як клас інструментів. Контрольований словник — це організований і стандартизований набір заздалегідь визначених термінів (слів та словосполучень), що використовуються для індексування контенту. Ключова функція такого словника полягає в тому, що він об'єднує варіантні терміни й синоніми одного поняття, зокрема різними мовами, установлює логічні зв'язки між поняттями та організовує їх у категорії. Тим самим забезпечується узгоджений спосіб опису й пошуку даних: усі учасники системи використовують одне й те саме позначення для одного й того самого явища.

Контрольовані словники реалізуються в кількох формах різного ступеня формалізації — від простих списків предметних рубрик та авторитетних файлів до тезаурусів, таксономій і буквено-цифрових класифікаційних схем. Найвищий рівень формалізації представлений онтологіями, які на базі контрольованих словників задають специфікацію концептуальної моделі предметної області: поняття та їх категорії, властивості, зв'язки, функції, обмеження й аксіоми. Словники можуть бути як загальнонауковими, так і специфічними для окремої дисципліни. Вибір форми не є питанням смаку: він визначається завданнями конкретного репозитарію, обсягом описуваного матеріалу та вимогами до глибини машинного оброблення метаданих.

Функції контрольованих словників у репозитаріях даних. Систематизація міжнародної практики дає підстави виокремити шість функцій, які контрольовані словники виконують у семантичній архітектурі репозитаріїв. Перша — забезпечення якості та узгодженості описів: стандартизовані терміни унеможливають використання різних назв для одного поняття, що є базовою умовою зіставності наборів даних. Друга — покращення пошуку та виявлення даних: зв'язки між термінами дають змогу розширювати або звужувати запит, підвищуючи водночас повноту і релевантність

результатів, причому цей ефект підтверджено емпірично (Gross et al., 2015). Третя — забезпечення інтероперабельності: спільні словники дозволяють різним системам «розуміти» одна одну, що спрощує обмін даними та інтеграцію між платформами, насамперед із DataCite, OpenAIRE і Європейською хмарою відкритої науки (EOSC).

Четверта функція — підтримання метаданих: структуровані словники є основою для створення стандартизованих описів змісту, контексту й характеристик наборів даних, причому потреба в такій підтримці простягається аж до рівня окремих змінних (Saldanha Bach Estevao & Klas, 2024). П'ята — довготривале збереження: стабільність термінології в часі критично важлива для доступності наукових даних у перспективі десятиліть, що прямо кореспондує з еталонною моделлю відкритої архівної інформаційної системи (ISO 14721:2025) та вимогами до сертифікації надійних цифрових репозитаріїв (ISO 16363:2025). Шоста — підтримання автоматизації: стандартизовані терміни дають змогу автоматизувати класифікування, індексування та оброблення даних, що є єдиним реалістичним способом управління великими масивами інформації. Дослідження С. Foster та колег (Foster et al., 2024) водночас показує, що автоматизація покриває не весь масив: навіть за наявності якісного перехресного показника від чверті до сорока відсотків описів потребують експертного втручання, тож словники не усувають ручну працю, а суттєво звужують її обсяг.

Показово, що перелічені функції прямо відповідають принципам FAIR: словники забезпечують знаходжуваність через якісне індексування, інтероперабельність через спільну семантику та придатність до повторного використання через повноту й однозначність описів. Водночас, як зауважують L. Vogt та співавтори (Vogt et al., 2025), самі принципи FAIR сформульовані надто загально щодо семантичного шару, тому контрольовані словники слід розглядати не як один із засобів дотримання FAIR, а як необхідну умову їх змістовної реалізації.

База нормативних документів (НД). Робота з контрольованими словниками спирається на розвинений корпус міжнародних стандартів, де основоположними НД щодо тезаурусів та їх сумісності з іншими словниками є ISO 25964-1:2011 «Information and documentation — Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval» та ISO 25964-2:2013 «Information and documentation — Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies», натомість ISO 1087:2019 «Terminology work and terminology science — Vocabulary» та ISO 5127:2017 «Information and documentation — Foundation and vocabulary» визначають термінологічний апарат галузі, а стандарти ISO 12620-1:2022 «Management of terminology resources — Data categories. Part 1: Specifications» та ISO 12620-2:2022 «Management of terminology resources — Data categories. Part 2: Repositories» регулюють управління термінологічними ресурсами і категоріями даних. Окрім того, на рівні метаданих ключовими НД є ISO 15836-1:2017 «Management of terminology resources — Data categories. Part 1: Specifications» та ISO 15836-2:2019 «Management of terminology resources — Data categories. Part 2: DCMI Properties and classes» (набір елементів метаданих Dublin Core), тоді як на рівні форматів обміну — специфікації W3C, насамперед SKOS і RDF Schema, які забезпечують машиночитане подання словникових структур. Практичні орієнтири задають рекомендації міжнародних організацій: COAR Controlled Vocabularies for Repositories разом із відповідним посібником з імплементації, реєстр BARTOC, матеріали Research Data Alliance та Infoterm, Рекомендація Європейської Комісії (ЄС) 2018/790 щодо доступу до наукової інформації та її збереження, а також Рекомендація ЮНЕСКО щодо відкритої науки 2021 р.

Міжнародний досвід галузевих словників. Світова практика напрацювала розгалужену систему галузевих словників, які фактично виконують роль де-факто стандартів опису у своїх дисциплінах: MeSH у медицині та біомедицині, AGROVOC у сільському господарстві, ChEBI у хімії й біохімії, Gene Ontology у біології та генетиці, MSC у математиці, ACM Computing Classification System у комп'ютерних науках, Ei Thesaurus в інженерії, NASA Thesaurus в аерокосмічній галузі, DDI Controlled Vocabularies у соціальних, поведінкових науках та економіці, HASSET у гуманітаристиці,

APA Thesaurus у психології, Getty Vocabularies (AAT, TGN, ULAN) у мистецтвознавстві та сфері культурної спадщини. Окремий клас становлять універсальні й авторитетні ресурси: Універсальна десяткова класифікація як багатомовна система для всіх галузей знань, Library of Congress Subject Headings як найширший словник предметних рубрик, GeoNames як глобальна база географічних назв і VIAF як міжнародний авторитетний файл осіб.

Для НРДД принциповим є те, що ці ресурси не потребують створення з нуля. Рациональна стратегія полягає в інтеграції наявних міжнародних словників з перекладом термінів українською мовою — і паралельно в інтеграції україномовної термінології до самих міжнародних словників, що є умовою повноцінної присутності українських знань у глобальному науковому обігу. Досвід, описаний A. Roy і S. Ghosh (Roy & Ghosh, 2025), застерігає при цьому від половинчастих рішень: якщо репозитарії впроваджують словники вибірково й неузгоджено, сумарний ефект для інтегрованих виявляється незначним.

Складові семантичної інфраструктури НРДД. Практична реалізація описаного підходу передбачає кілька взаємопов'язаних напрямів роботи. Перший — вивчення міжнародного досвіду та відбір словників для інтеграції з подальшим перекладом українською мовою. Другий — розроблення архітектури системи контрольованих словників зі схемами даних і форматами обміну (XML, JSON, RDF, SKOS), яка визначає структуру, взаємозв'язки та атрибути кожного терміна й гарантує легкість експорту та імпорту. Третій — інтеграція україномовної термінології до чинних міжнародних словників. Четвертий — стандартизація класифікацій задля єдиної системи опису дослідницьких даних. П'ятий — розроблення програмного інтерфейсу (API), що надасть програмний доступ до даних словників і забезпечить єдину валідацію метаданих. Шостий — створення вебінтерфейсу для перегляду, пошуку та експорту даних у стандартних форматах. Сьомий — підготовка посібника користувача та методичних рекомендацій для дослідників із чіткими інструкціями та практичними прикладами.

Результати опитування Y. Zhang та колег (Zhang et al., 2015) дають підстави уточнити вимоги до п'ятого й шостого напрямів: користувачі очікують не доступу до одного словника, а можливості працювати з кількома словниками одночасно, що робить багатословникову архітектуру не додатковою опцією, а вихідною вимогою до проектування.

Ризики та обмеження. Побудова багатомовної семантичної інфраструктури не є суто інженерним завданням. Вона потребує залучення фахівців зі стандартизації термінології, лінгвістики та перекладу — лише за цієї умови можна забезпечити точність відтворення наукових понять і їх відповідність міжнародним практикам. Механічний переклад термінологічних систем неминує породжує викривлення, які згодом закріплюються в метаданих і відтворюються в усіх похідних описах. Другий ризик пов'язаний із версійністю: словники оновлюються, терміни змінюють статус, тому система має підтримувати стійкі ідентифікатори понять і фіксувати історію змін, інакше описи, створені в різні роки, стануть незіставними. Третій ризик — інституційний: підтримання словників потребує сталого фінансування та закріплення відповідальності, і саме відсутність такого закріплення найчастіше перетворює якісно спроектовані системи на застарілі протягом кількох років. Нарешті, узгодженість опису слід розглядати в ширшому контексті цілісності дослідницьких даних, що охоплює також планування збору, визначення змінних та відтворюваність оброблення (Miller & Spiegel, 2025).

ВИСНОВКИ. Контрольовані словники є не допоміжним сервісом, а базовим елементом семантичної архітектури репозитаріїв даних, від якого залежать якість метаданих, ефективність пошуку, інтегрованих із міжнародними інфраструктурами та можливість довготривалого збереження наукової інформації. Емпіричні дані свідчать, що відмова від контрольованої лексики призводить до втрати значної частки релевантних результатів пошуку, а автоматизоване застосування словників дає змогу стандартизувати більшість описів, істотно скорочуючи обсяг ручної праці.

Для Національного репозитарію дослідницьких даних розроблення та впровадження системи контрольованих словників з інтеграцією міжнародних ресурсів і адаптацією їх для української мови є передумовою входження українських наукових даних до глобального дослідницького простору й реалізації положень Національного плану щодо відкритої науки та Дорожньої карти інтеграції України до Європейського дослідницького простору. Запропонований склад семантичної інфраструктури охоплює сім взаємопов'язаних напрямів: відбір і переклад словників, розроблення архітектури та форматів обміну, інтеграцію україномовної термінології до міжнародних ресурсів, стандартизацію класифікацій, розроблення API, створення вебінтерфейсу і методичне забезпечення.

Перспективи подальших досліджень пов'язані з розробленням методики оцінювання якості перекладу термінологічних систем, з експериментальною перевіркою ефективності багатословникового пошуку на масиві українських наукових даних, а також з опрацюванням моделі сталого інституційного підтримання словників. Постійне оновлення контрольованих словників має розглядатися як тривала інституційна функція, а не як разовий проєкт.

ПОДЯКИ. Публікацію здійснено в межах виконання пункту «Створення Національного репозитарію дослідницьких даних» із затвердженого 17.06.2026 р. Кабінетом міністрів України плану дій із впровадження Ініціативи «Партнерство “Відкритий Уряд” у 2026–2027 роках», а також в межах виконання НТР «Дослідження міжнародного досвіду управління науковими даними та розробка методичних засад створення Національного репозитарію наукових даних» (Державний реєстраційний номер НДДКР: 0125U001107, виконавець ДНТБ України, фінансування МОН України).

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

- Basic Register of Thesauri, Ontologies & Classifications (BARTOC). (n.d.). Vocabularies. <https://bartoc.org/vocabularies>
- Cabinet of Ministers of Ukraine. (2022, October 8). Pro zatverdzhennia Natsionalnoho planu shchodo vidkrytoi nauky [On approval of the National Plan for Open Science] (Order No. 892-r). Zakonodavstvo Ukrainy. <https://zakon.rada.gov.ua/laws/show/892-2022-%D1%80> [in Ukrainian]
- Cabinet of Ministers of Ukraine. (2026). Pro zatverdzhennia planu dii iz vprovadzhennia Initsiatyvy “Partnerstvo “Vidkryti Uriad” u 2026-2027 rokakh [On the Approval of the Action Plan for the Implementation of the “Open Government Partnership” Initiative in 2026–2027] (Order No. 603-p). <https://zakon.rada.gov.ua/laws/show/603-2026-%d1%80#Text> [in Ukrainian]
- Confederation of Open Access Repositories. (n.d.). COAR controlled vocabularies for repositories. <https://vocabularies.coar-repositories.org/>
- European Commission. (2018, April 25). Commission Recommendation (EU) 2018/790 on access to and preservation of scientific information. Official Journal of the European Union, L 134, 12–18. <http://data.europa.eu/eli/reco/2018/790/oj>
- Foster, C., Wignall, J., Kovach, S., Choksi, N., Allen, D., Trgovcich, J., Rochester, J. R., Ceger, P., Daniel, A., Hamm, J., Truax, J., Blake, B., McIntyre, B., Sutherland, V., Stout, M. D., & Kleinstreuer, N. (2024). Standardizing extracted data using automated application of controlled vocabularies. *Environmental Health Perspectives*, 132(2), 027006. <https://doi.org/10.1289/EHP13215>
- Gross, T., Taylor, A. G., & Joudrey, D. N. (2015). Still a lot to lose: The role of controlled vocabulary in keyword searching. *Cataloging & Classification Quarterly*, 53(1), 1–39. <https://doi.org/10.1080/01639374.2014.917447>
- International Organization for Standardization. (2017). Information and documentation – Foundation and vocabulary (ISO Standard No. ISO 5127:2017). <https://www.iso.org/ru/standard/59743.html>
- International Organization for Standardization. (2017). Information and documentation – The Dublin Core metadata element set. Part 1: Core elements (ISO Standard No. ISO 15836-1:2017). <https://www.iso.org/ru/standard/71339.html>

- International Organization for Standardization. (2019). Information and documentation – The Dublin Core metadata element set. Part 2: DCMI Properties and classes (ISO Standard No. ISO 15836-2:2019). <https://www.iso.org/standard/71341.html>
- International Organization for Standardization. (2011). Information and documentation – Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval (ISO Standard No. ISO 25964-1:2011). <https://www.iso.org/ru/standard/53657.html>
- International Organization for Standardization. (2013). Information and documentation – Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies (ISO Standard No. ISO 25964-2:2013). <https://www.iso.org/ru/standard/53658.html>
- International Organization for Standardization. (2022). Management of terminology resources – Data categories. Part 1: Specifications (ISO Standard No. ISO 12620-1:2022). <https://www.iso.org/standard/79078.html>
- International Organization for Standardization. (2022). Management of terminology resources – Data categories. Part 2: Repositories (ISO Standard No. ISO 12620-2:2022). <https://www.iso.org/ru/standard/79018.html>
- International Organization for Standardization. (2019). Terminology work and terminology science – Vocabulary (ISO Standard No. ISO 1087:2019). <https://www.iso.org/standard/62330.html>
- Miller, G., & Spiegel, E. (2025). Guidelines for research data integrity (GRDI). *Scientific Data*, 12(1), 95. <https://doi.org/10.1038/s41597-024-04312-x>
- Ministry of Education and Science of Ukraine. (2026, January 2) Zatverdzheno Dorozhniu kartu intehtatsii Ukrainy do Yevropeiskoho doslidnytskoho prostoru do 2027 roku [Roadmap for Ukraine's Integration into the European Research Area by 2027 Approved] [News]. https://mon.gov.ua/news/zatverdzheno-dorozhniu-kartu-intehtatsii-ukrainy-do-ievropeiskoho-doslidnytskoho-prostoru-do-2027-roku?__cf_chl_tk=xrTijptBfkFL4M2G6YB5n_AR5pibOn6YGXNzLTYB8DY-1789040372-1.0.1.1-nzUX2zmE Zp2gY6Vg83eF6gX27fatKV9zIH90.JFUy0w [in Ukrainian]
- Roy, A., & Ghosh, S. (2025). From chaos to clarity: How controlled vocabularies shape Indian social science repositories. *The Electronic Library*, 43(2), 154–173. <https://doi.org/10.1108/EL-09-2024-0277>
- Saldanha Bach Estevao, J., & Klas, C.-P. (2024). Enhancing FAIR compliance: A controlled vocabulary for mapping Social Sciences survey variables. *IASSIST Quarterly*, 48(2). <https://doi.org/10.29173/iq1118>
- Smith, C. (2021). Controlled vocabularies: Past, present and future of subject access. *Cataloging & Classification Quarterly*, 59(2–3), 186–202. <https://doi.org/10.1080/01639374.2021.1881007>
- UNESCO. (2021). UNESCO Recommendation on Open Science. Paris: United Nations Educational, Scientific and Cultural Organization.
- Verkhovna Rada of Ukraine (2026). Pro vnesennia zmin do Zakonu Ukrainy “Pro naukovu i naukovotekhnichnu diialnist” shchodo pytan doslidnytskoi infrastruktury ta pidtrymky molodykh vchenykh [On Amendments to the Law of Ukraine “On Scientific and Scientific-Technical Activities” Concerning Research Infrastructure and Support for Young Scientists] (Law of Ukraine of February 25, 2026, No. 4794-IX). *Zakonodavstvo Ukrainy* [Legislation of Ukraine]. <https://zakon.rada.gov.ua/laws/show/4794-20#Text> [in Ukrainian].
- Vogt, L., Strömert, P., Matentzoglou, N., Karam, N., Konrad, M., Prinz, M., & Baum, R. (2025). Suggestions for extending the FAIR Principles based on a linguistic perspective on semantic interoperability. *Scientific Data*, 12(1), 688. <https://doi.org/10.1038/s41597-025-05011-x>
- Zhang, Y., Ogletree, A., Greenberg, J., & Rowell, C. (2015). Controlled vocabularies for scientific data: Users and desired functionalities. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–8. <https://doi.org/10.1002/pra2.2015.145052010054>

Petrenko Oksana,
position, academic degree, academic title
State Scientific and Technical Library of Ukraine, Kyiv, Ukraine,
ORCID: 0000-0001-5788-8192, e-mail: lo.petrenko@dntb.gov.ua

Khymenko Bohdan,
Researcher, PhD student
State Scientific and Technical Library of Ukraine, Kyiv, Ukraine,
ORCID: 0009-0007-2612-8909, e-mail: bogkhym@gmail.com

THE ROLE OF CONTROLLED VOCABULARIES FOR RESEARCH DATA REPOSITORIES

ABSTRACT. *The purpose of the study is to substantiate the role of controlled vocabularies as a core element of the semantic architecture of research data repositories and to define the set of tasks required to build such a system for a national repository. The study applies documentary and comparative analysis of international standards and domain vocabularies, content analysis of recommendations issued by international organisations, terminological analysis and functional modelling. Six principal functions of controlled vocabularies in repositories are established: ensuring the quality and consistency of descriptions, improving search and data discovery, supporting interoperability, structuring metadata, enabling long-term preservation and supporting automated processing. The normative framework is systematised and the leading domain vocabularies suitable for integration are characterised. A composition of the semantic infrastructure of a national repository is proposed, covering vocabulary architecture, exchange formats, an application programming interface, a web interface and methodological support. It is concluded that vocabulary maintenance must be a permanent institutional function rather than a one-off project.*

KEYWORDS: *controlled vocabularies; research data repository; metadata; interoperability; FAIR principles; open science.*