

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХЕРСОНСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
Факультет комп'ютерних наук, фізики та математики
Кафедра комп'ютерних наук та програмної інженерії

РОЗРОБКА МОДЕЛЕЙ ДАНИХ ТА ВИМОГ ДО
ІНФОРМАЦІЙНОЇ СИСТЕМИ ОБЛІКУ НАУКОМЕТРИЧНОЇ
АКТИВНОСТІ

Кваліфікаційна робота (проект)
на здобуття ступеня вищої освіти бакалавр

Виконав: здобувач 4 курсу 12-461 групи
Спеціальності: 126 Інформаційні системи
та технології

Освітньо-професійної програми:
«Інформаційні системи та технології»
Зубков Данііл Ігорович

Керівник: кандидат педагогічних наук,
доцент, Вінник Максим Олександрович

Рецензент: професор кафедри
інформатики і кібернетики
Мелітопольського державного
педагогічного університету імені Богдана
Хмельницького Круглик Владислав
Сергійович

Івано-Франківськ – 2026

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	4
ВСТУП	5
РОЗДІЛ 1 ТЕОРЕТИЧНІ ЗАСАДИ ОБЛІКУ НАУКОМЕТРИЧНОЇ АКТИВНОСТІ ТА ОСОБЛИВОСТІ ЇЇ АВТОМАТИЗАЦІЇ.....	8
1.1 Наукометрія як галузь знань та інструмент оцінювання наукової діяльності.....	8
1.2 Основні наукометричні показники, платформи та джерела даних	10
1.3 Особливості наукометричної активності як об’єкта інформаційного обліку	15
1.4 Проблеми та перспективи автоматизації обліку наукометричної активності	19
Висновки до розділу 1	23
РОЗДІЛ 2 ФОРМУВАННЯ ВИМОГ ДО ІНФОРМАЦІЙНОЇ СИСТЕМИ ОБЛІКУ НАУКОМЕТРИЧНОЇ АКТИВНОСТІ	27
2.1 Аналіз практичного контексту та досвіду використання системи publication.ksru.edu	27
2.2 Обґрунтування напрямів розвитку інформаційної системи обліку наукометричної активності	30
2.3 Формування функціональних вимог до системи	36
2.4 Формування нефункціональних вимог до системи	42
Висновки до розділу 2	47
РОЗДІЛ 3 РОЗРОБКА МОДЕЛЕЙ ДАНИХ ІНФОРМАЦІЙНОЇ СИСТЕМИ ОБЛІКУ НАУКОМЕТРИЧНОЇ АКТИВНОСТІ	50
3.1 Визначення основних сутностей предметної області	50

3.2 Розробка концептуальної моделі даних.....	55
3.3 Розробка логічної моделі даних	61
3.4 Узагальнення критеріїв представлення наукометричних показників у структурі даних системи	70
Висновки до розділу 3	75
ВИСНОВКИ.....	77
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	82
ДОДАТКИ.....	89
Додаток А.....	89
Додаток Б	91
Додаток В.....	93

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

БД — база даних;

ЗВО — заклад вищої освіти;

ІС — інформаційна система;

МОН України — Міністерство освіти і науки України;

НПП — науково-педагогічні працівники;

ХДУ — Херсонський державний університет;

API (Application Programming Interface) — прикладний програмний інтерфейс;

DOI (Digital Object Identifier) — цифровий ідентифікатор об'єкта;

ER (Entity-Relationship) — модель «сутність — зв'язок»;

FK (Foreign Key) — зовнішній ключ у реляційній базі даних;

JSON (JavaScript Object Notation) — текстовий формат обміну даними;

ORCID (Open Researcher and Contributor ID) — відкритий цифровий ідентифікатор дослідника та автора;

PK (Primary Key) — первинний ключ у реляційній базі даних;

REST (Representational State Transfer) — архітектурний стиль взаємодії компонентів розподіленого застосунку;

ROR (Research Organization Registry) — відкритий реєстр ідентифікаторів дослідницьких організацій;

UML (Unified Modeling Language) — уніфікована мова моделювання;

URL (Uniform Resource Locator) — уніфікований покажчик ресурсу.

Актуальність теми. У сучасному університетському середовищі відомості про публікаційну активність, цитування та цифрові ідентифікатори дослідників формуються одночасно в кількох різномірних інформаційних середовищах. Це обумовлює складність їхнього узгодженого обліку, систематизації та використання для внутрішнього моніторингу і прийняття управлінських рішень [1; 2].

У Херсонському державному університеті з 2020 року функціонує електронний сервіс рейтингування *publication.kspu.edu*, запроваджений наказом ректора від 23.04.2020 № 385-Д [3]. Сервіс розроблено колективом авторів — М. О. Вінником, М. Ю. Полторацьким та О. В. Співаковським; у його межах реалізовано багатоплатформний моніторинг показників, формування персональних сторінок науковців та аналітичних представлень результативності підрозділів [4–6]. Водночас подальший розвиток наявного рішення потребує уточнення й формалізації моделі даних, системної підтримки історичних зрізів, уніфікованого представлення різномірних наукометричних показників, фіксації джерел походження та статусів верифікації даних, а також розширення інтеграційних можливостей для роботи з новими платформами. Це створює прикладну потребу у розробці моделей даних та вимог до інформаційної системи обліку наукометричної активності з позицій спеціальності 126 «Інформаційні системи та технології», стандарт вищої освіти за якою затверджено наказом МОН України від 12.12.2018 № 1380 [8].

Стан розробленості проблеми. Теоретичні засади наукометрії закладені у працях Д. де Солла Прайса, Ю. Гарфілда та В. В. Налімова. Сучасні підходи до оцінювання наукової продуктивності, зокрема h-індекс Дж. Хірша, інтегровані у провідні платформи (Scopus, Web of Science, Google Scholar). Принципи відповідального використання

метрик закріплено в Лейденському маніфесті та декларації DORA [9–16]. Водночас прикладні аспекти моделювання структур даних для інтеграції багатоплатформних показників з урахуванням еволюції API зовнішніх сервісів (Crossref, OpenAlex, ORCID) в університетських інформаційних системах потребують подальшого опрацювання, зокрема в частині уніфікації сутностей предметної області для забезпечення цілісності та перевірюваності даних.

Мета роботи — розробити моделі даних та сформувані вимоги до вдосконалення інформаційної системи обліку наукометричної активності на основі аналізу функціонування сервісу *publication.kspu.edu* з урахуванням потреб багатоплатформного моніторингу, накопичення історичних даних і розширення переліку показників.

Завдання дослідження:

дослідити наукометрію як предметну область інформаційної системи, схарактеризувати основні наукометричні платформи та визначити особливості наукометричної активності як об'єкта цифрового обліку;

проаналізувати практичний контекст функціонування *publication.kspu.edu*, виявити функціональні та інформаційні обмеження поточної реалізації й обґрунтувати напрями розвитку системи;

сформувані функціональні вимоги до інформаційної системи обліку наукометричної активності;

сформувані нефункціональні вимоги до системи з урахуванням моделей якості програмного забезпечення ISO/IEC 25010 та якості даних ISO/IEC 25012;

визначити основні сутності предметної області та розробити концептуальну і логічну моделі даних;

обґрунтувати підхід до представлення наукометричних показників у структурі ІС та визначити перспективний перелік платформ і метрик для підтримки.

Об'єкт дослідження — процеси обліку, накопичення та аналітичного опрацювання наукометричної активності в університетському цифровому середовищі.

Предмет дослідження — моделі даних і вимоги до інформаційної системи обліку наукометричної активності, орієнтованої на багатоплатформний моніторинг та збереження історичних зрізів.

Методи дослідження: *аналіз і узагальнення фахової літератури* — для дослідження наукометрії як предметної області; *системний аналіз* — для дослідження функціонування *publication.kspu.edu*; *методи збору та аналізу вимог* — для формування функціональних вимог; *використання стандартів ISO/IEC 25010 та ISO/IEC 25012* — для обґрунтування і структуризації нефункціональних вимог; *методи інформаційного моделювання (ER-моделювання)* — для побудови моделей даних; *класифікація та структуризація* — для впорядкування платформ і показників.

Практичне значення одержаних результатів полягає в можливості застосування сформованих вимог і запропонованих моделей даних як основи для подальшого проєктування та розвитку цифрової інфраструктури університету: розробки архітектури бази даних, специфікації серверної частини та інтеграції нових моніторингових платформ [3–6].

Структура роботи. Робота складається зі вступу, трьох розділів, висновків, списку використаних джерел і додатків. Загальний обсяг основного тексту — ____ сторінок, список використаних джерел містить ____ найменувань.

РОЗДІЛ 1 ТЕОРЕТИЧНІ ЗАСАДИ ОБЛІКУ НАУКОМЕТРИЧНОЇ АКТИВНОСТІ ТА ОСОБЛИВОСТІ ЇЇ АВТОМАТИЗАЦІЇ

1.1 Наукометрія як галузь знань та інструмент оцінювання наукової діяльності

Наукометрія є напрямом досліджень, що зосереджується на кількісному аналізі науки, наукової комунікації, динаміки публікацій і цитування, наукової продуктивності та взаємозв'язків між авторами, установами й джерелами інформації. У фаховій літературі її розглядають як споріднену з бібліометрією та інформетрією сферу, з виразним акцентом на вивченні науки як соціального та інформаційного феномена [1; 2].

Історичне становлення наукометрії пов'язане з розвитком «науки про науку» та з появою інструментів кількісного аналізу наукової літератури. Однією з ключових постатей у формуванні цього напрямку є Дерек де Солла Прайс, який обґрунтував закономірності зростання, структуризації та самоорганізації наукового знання [10]. Фундаментальну роль відіграли праці Юджина Гарфілда, який обґрунтував значення індексування цитувань як інструмента виявлення зв'язків між науковими роботами та заклав підвалини цитатного аналізу [11]. Сам термін *scientometrics* пов'язують із працями В. В. Налімова та подальшим інституційним оформленням галузі через створення профільного журналу *Scientometrics*, що спеціалізується на кількісних аспектах науки і наукової комунікації [12].

Предмет наукометрії значно ширший за механічний підрахунок публікацій чи цитувань. Він охоплює аналіз продуктивності науковців та установ, вивчення комунікаційних зв'язків між дослідниками, виявлення тематичних кластерів, оцінювання видимості наукових результатів у міжнародному просторі, а також дослідження впливу форм

представлення результатів на наукову репутацію та управлінські рішення [9]. Розвиток галузі нерозривно пов'язаний з еволюцією інформаційної інфраструктури науки: якщо на ранніх етапах увага зосереджувалася на публікаціях і цитатних зв'язках, то нині в центрі опиняються цифрові профілі дослідників, стійкі ідентифікатори (зокрема ORCID), інституційні афіліації, профілі журналів та різні типи дослідницьких результатів [13].

У прикладному вимірі наукометрія є одним з основних інструментів оцінювання наукової діяльності. На її основі формуються індикатори, що використовуються для аналізу результативності окремих науковців, колективів, структурних підрозділів, журналів та установ. До найвідоміших належать кількість публікацій, кількість цитувань, h-індекс, запропонований Дж. Хіршем як узагальнений показник, що поєднує продуктивність дослідника та цитованість його праць [14], а також показники рівня журналу — CiteScore, SNIP, SJR, Journal Impact Factor.

Водночас жоден окремих показник не здатний повноцінно відобразити якість наукової діяльності. У сучасній міжнародній дискусії ця думка є консенсусною. Лейденський маніфест прямо наголошує, що кількісне оцінювання має підтримувати, а не підміняти експертне, а самі метрики потрібно застосовувати з урахуванням контексту дослідження та галузевих відмінностей [15]. Аналогічний акцент міститься в декларації DORA, яка застерігає від спрощеного використання окремих метрик як універсального еквівалента наукової якості [16]. Для академічного середовища це означає, що наукометрія має розглядатися не як механізм автоматичного винесення висновків, а як інструмент підтримки аргументованого аналізу.

Для теми цієї роботи принципово важливим є те, що наукометрія формує специфічну предметну область для інформаційного моделювання. У межах цієї області взаємодіють різні класи об'єктів:

автори, публікації, наукові профілі, цитування, індикатори, афіліації, часові зрізи, джерела даних і зв'язки між авторами та установами. Частина цих об'єктів є відносно стабільною (відомості про автора чи видання), а частина — динамічною (значення показників, цитувань, профілів дослідника на різних платформах). Крім того, одна і та сама характеристика може по-різному подаватися в різних системах. Отже, наукометрія є складною, багатокomпонентною і структурно неоднорідною предметною областю, що потребує системного підходу до формування вимог і розробки моделей даних інформаційної системи. Саме з цієї позиції наукометрія виступає теоретичною основою подальших розділів роботи.

1.2 Основні наукометричні показники, платформи та джерела даних

Для побудови інформаційної системи обліку наукометричної активності необхідно визначити, які саме дані формуються у різних цифрових середовищах, до яких об'єктів вони належать і чому ці дані не можуть розглядатися як однорідні. Сучасне наукове середовище формує відомості про результати дослідницької діяльності у п'яти взаємопов'язаних класах інформаційних систем: міжнародних реферативно-цитатних базах, академічних пошукових платформах, системах ідентифікації дослідників, інфраструктурах відкритих метаданих та аналітичних сервісах [1; 2; 9; 13]. Джерела даних відрізняються не лише за назвою, а й за функціональною роллю, складом сутностей, доступністю API та типами підтримуваних показників, що принципово важливо для проектування системи.

До міжнародних реферативно-цитатних баз належать насамперед Scopus і Web of Science. Scopus позиціонується як реферативна та цитатна база наукової літератури з метриками на рівні авторів, публікацій, джерел та установ; до підтримуваних показників належать

кількість цитувань, h-індекс, журнальні метрики CiteScore, SNIP і SJR [17; 18; 19]. Web of Science функціонує як аналітична платформа пошуку наукової інформації та підтримує роботу з профілями дослідників, публікаціями, цитуваннями та журнальними метриками Journal Citation Reports [20; 21; 22]. Уже в межах двох міжнародних платформ дані мають різну природу: одна частина описує автора, інша — окрему роботу, ще інша — джерело публікації або інституційний рівень.

До академічних пошукових платформ належить Google Scholar — попри те, що формально він не є реферативно-цитатною базою у класичному розумінні. У межах Google Scholar Citations система обчислює загальну кількість цитувань, h-індекс та i10-індекс автора, а Google Scholar Metrics використовує h5-індекс і h5-медіану для оцінювання видань за п'ятирічним вікном публікацій [23; 24]. Цей приклад показовий тим, що одна платформа одночасно надає показники авторського рівня та рівня джерел наукової комунікації — отже, у моделі даних потрібне явне розрізнення об'єкта застосування метрики.

До систем стійкої цифрової ідентифікації дослідників належить ORCID — безкоштовний унікальний постійний ідентифікатор для осіб, залучених до досліджень [13; 25]. ORCID не є наукометричною базою у класичному розумінні, оскільки його основна функція — не розрахунок цитатних індикаторів, а однозначне пов'язування дослідника з його працями, афіліаціями, фінансуванням і рецензуванням. Для системи обліку наукометричної активності це означає, що поряд із бібліографічними та цитатними даними необхідно моделювати ідентифікаційні сутності, які забезпечують коректне зіставлення інформації між різними платформами. Технічну роботу з ORCID забезпечують публічний API та документація розробника [53].

До інфраструктур відкритих метаданих належать Crossref та OpenAlex. Crossref через відкритий REST API надає бібліографічні метадані з пов'язаними атрибутами — DOI, ORCID, ROR, дані про

фінансування, ліцензії та анотації [26; 27; 35]. OpenAlex позиціонується як відкритий каталог глобальної дослідницької системи, що поєднує роботи (works), авторів (authors), інституції (institutions), джерела (sources) і тематики у зв'язаній структурі даних [28; 33; 51; 54]. Ці ресурси демонструють інший підхід до побудови системи: не через фіксований перелік наперед заданих полів, а через взаємопов'язані сутності, з яких можуть бути сформовані або уточнені аналітичні показники. Саме така архітектура є орієнтиром для розвитку моделі даних publication.kspu.edu.

До аналітичних сервісів зі зв'язаною структурою належить Dimensions, що репрезентується як велика база дослідницьких даних, доступна через вебплатформу та прикладні сервіси [29; 30; 31]. У межах цієї платформи можна отримувати як цитатні, так і альтернативні метрики, працювати з профілями дослідників, установ і глобальними дослідницькими трендами. Межі між реферативною базою, аналітичною платформою та профільним сервісом у сучасному науковому середовищі поступово стираються, тому майбутня система обліку наукометричної активності повинна бути готовою до інтеграції даних різної природи: ідентифікаційних, бібліографічних, цитатних, аналітичних і контекстних. Узагальнену порівняльну характеристику основних платформ і інфраструктур як джерел даних подано в таблиці 1.1.

Таблиця 1.1

Порівняльна характеристика основних джерел наукометричних та бібліографічних даних

Платформа / інфраструктура	Основне призначення	Доступність даних (API)	Ключові сутності для обліку
Scopus (Elsevier)	Реферативно-цитатна база, джерело метрик	Обмежений комерційний API	Публікації, профілі авторів, цитування, журнальні метрики (CiteScore, SNIP, SJR)

Платформа / інфраструктура	Основне призначення	Доступність даних (API)	Ключові сутності для обліку
Web of Science (Clarivate)	Цитатна аналітична платформа	Обмежений комерційний API	Публікації, профілі дослідників, цитування, JCR-показники
Google Scholar	Академічна пошукова платформа	Офіційний публічний API не надається	Профілі авторів, цитування, h-індекс, i10-індекс, h5-показники
ORCID	Система стійкої ідентифікації дослідників	Відкритий публічний REST API	Ідентифікатор особи, афіліації, зв'язки з роботами
Crossref	Відкрита інфраструктура метаданих	Відкритий REST API (JSON)	DOI, бібліографічні метадані, ORCID, ROR, фінансування
OpenAlex	Відкритий каталог глобальної дослідницької системи	Відкритий API / дампи даних	Works, authors, institutions, sources, topics
Dimensions	Зв'язана аналітична база дослідницьких даних	Вебплатформа та прикладні сервіси	Публікації, профілі дослідників, цитатні й альтернативні метрики

Якщо перейти від платформ до самих показників, то базові наукометричні індикатори доцільно класифікувати за рівнем застосування: показники авторського рівня (кількість публікацій, кількість цитувань, h-індекс, i10-індекс), показники рівня публікації (цитованість конкретної роботи, альтметрики), показники рівня джерела публікації (CiteScore, SJR, SNIP, Journal Impact Factor, h5-індекс) та інституційні показники (агреговані рейтинги, мережеві характеристики співавторства) [14; 17; 18; 23; 24]. Кількість публікацій відображає продуктивність автора або установи, але сама по собі не характеризує

вплив роботи. Кількість цитувань залежить від галузі, віку публікації та політики індексування платформи. *h*-індекс, запропонований Дж. Хіршем, поєднує продуктивність і цитованість в один узагальнений показник [14], а *i10*-індекс показує кількість праць, процитованих щонайменше десять разів [23]. Журнальні метрики стосуються джерела публікації, а не автора, тому не можуть інтерпретуватися як характеристика індивідуальної наукової результативності [17; 22] — це принциповий поділ для розробки моделі даних, оскільки запобігає змішуванню різнорівневих сутностей в одній площині обліку.

Один і той самий дослідник може мати різні значення кількості публікацій, цитувань або *h*-індексу у Scopus, Web of Science та Google Scholar через відмінності в охопленні джерел, механізмах індексування, частоті оновлення даних та логіці зіставлення автора з роботами [17; 23; 34]. Для проєктування інформаційної системи це означає, що значення показника без вказівки джерела втрачає частину свого змісту — система повинна фіксувати не лише саме значення, а й платформу його отримання, дату фіксації, об'єкт застосування та, за потреби, спосіб одержання.

Джерелом даних для системи можуть бути не лише зовнішні платформи, а й внутрішні університетські ресурси: інституційні репозитарії, внутрішні профілі науково-педагогічних працівників, бази даних публікацій, кадрові та структурні довідники. Зовнішні сервіси дають відомості про міжнародну представленість, цитованість і профільні ідентифікатори, а внутрішні джерела — про посаду, кафедру, факультет, належність автора до установи та період його активності [13; 36; 41]. Повноцінний облік наукометричної активності має гібридний характер і потребує інтеграції зовнішніх і внутрішніх джерел даних у межах єдиного інформаційного середовища.

З аналізу платформ і показників випливає висновок, важливий для подальшої роботи: інформаційна система не повинна будуватися як

набір жорстко зафіксованих полів для окремих платформ або метрик, бо в такому разі поява нової платформи чи нового індикатора вимагатиме зміни структури сховища. Більш коректним є підхід, за якого різномірні показники описуються через спільний набір характеристик — назву, тип, рівень застосування, платформу-джерело, одиницю вимірювання, часову характеристику, значення та статус походження. Цей підхід деталізовано у підрозділі 3.4 у вигляді узагальнених критеріїв представлення показників. Аналіз основних наукометричних платформ, показників і джерел даних показав, що предметна область є різномірною за складом сутностей, рівнем агрегування та характером метрик; майбутня інформаційна система повинна проєктуватися як інтеграційне середовище, здатне працювати з різними типами джерел, підтримувати зв'язування ідентифікаторів, фіксувати контекст походження показників і забезпечувати масштабоване представлення різномірних метрик.

1.3 Особливості наукометричної активності як об'єкта інформаційного обліку

Наукометрична активність як об'єкт інформаційного обліку є значно складнішою, ніж це може здаватися при поверховому розгляді. У найзагальнішому розумінні її нерідко зводять до переліку публікацій, цитувань або індексів. Однак з позицій інформаційних систем ідеться не про окремі числові характеристики, а про багаторівневий масив взаємопов'язаних даних, що описує науковця, результати його діяльності, способи їх цифрового представлення та аналітичні параметри, які змінюються з часом. ORCID репрезентує дослідника через унікальний постійний ідентифікатор, тоді як Scopus Author Profiles подає стандартизований огляд наукового доробку автора, його афіліацій, цитувань і співавторств. Уже на цьому рівні очевидно, що облік наукометричної активності охоплює не один тип даних, а кілька різних

сутностей: особу, профіль, публікацію, організацію, зв'язки між ними та показники впливу [13; 32].

Однією з ключових рис наукометричної активності є її структурна неоднорідність. У межах одного дослідницького профілю можуть одночасно поєднуватися бібліографічні відомості про публікації, дані про авторів і співавторів, інформація про афіліації, цифрові ідентифікатори, тематичні зв'язки, джерела оприлюднення, а також числові індикатори, що відображають рівень цитованості чи видимості роботи. OpenAlex прямо подає глобальну дослідницьку систему як граф взаємопов'язаних сутностей, до яких належать роботи, автори, інституції, джерела, тематики та інші об'єкти. Через це наукометричну активність не можна коректно описати спрощеною моделлю «автор — список публікацій», адже сама предметна область має мережеву, а не лінійну структуру [28; 33].

Не менш важливою є наявність у цій предметній області як відносно стабільних, так і змінних даних. До стабільніших характеристик можна віднести назву публікації, DOI, тип документа або ORCID iD автора. Натомість кількість цитувань, h-index, склад авторського профілю, афіліації чи навіть перелік пов'язаних робіт можуть змінюватися з часом. Причинами таких змін є індексування нових джерел, уточнення метаданих, корекція профілів або оновлення самої платформи. Elsevier окремо підкреслює, що точність і актуальність авторського профілю мають критичне значення, а Scopus передбачає процедури виправлення профілів, афіліацій, цитувань і документів. Для інформаційної системи це означає, що фіксувати потрібно не тільки саме значення показника, а й час його отримання, джерело та контекст актуальності [13; 32; 34].

Ще однією важливою особливістю є залежність значення показника від платформи-джерела. Один і той самий дослідник у різних системах може мати різну кількість публікацій, різну кількість цитувань

і навіть різний h-index. Це зумовлено тим, що платформи відрізняються охопленням джерел, правилами індексування, частотою оновлення інформації та механізмами зіставлення авторів із роботами. Google Scholar, Scopus, Web of Science, OpenAlex і Dimensions працюють з різними масивами даних і спираються на різні аналітичні моделі. Через це будь-який показник без зазначення джерела втрачає частину свого наукового змісту. У системі обліку він має зберігатися разом із посиланням на платформу, датою фіксації, рівнем застосування та, за потреби, способом отримання [17; 20; 23; 24; 28; 29; 31].

Окремою проблемою є ідентифікація дослідника та зіставлення даних, отриманих із різних джерел. У науковому середовищі одна й та сама особа може бути представлена по-різному: через варіанти написання імені, зміну афіліації, розщеплення або об'єднання профілів, різні мовні форми прізвища, а також через неточності в метаданих. Саме тому ORCID розглядається як постійний ідентифікатор, покликаний забезпечити надійний зв'язок між дослідником та його внеском у науку. Crossref також прямо вказує, що в його REST API можуть бути доступні ORCID- і ROR-ідентифікатори, які підвищують якість зв'язування робіт, авторів та інституцій. Для інформаційної системи обліку наукометричної активності це означає необхідність передбачити не лише атрибути особи, а й механізми зберігання зовнішніх ідентифікаторів та зв'язків між різними профілями одного автора [13; 26].

Додаткову складність створює неоднорідність самих показників. Частина з них належить до авторського рівня, частина — до окремої публікації, частина — до журналу, конференції чи іншого джерела. Одні показники мають абсолютний характер, наприклад кількість цитувань, інші є узагальненими або похідними, як h-index чи CiteScore. Деякі з них платформа надає безпосередньо, інші можуть бути обчислені на основі відкритих метаданих або використані як допоміжні характеристики

аналітики. Тому в інформаційному обліку показник не може розглядатися як ізольоване числове поле. Він має бути пов'язаний із типом, об'єктом застосування, джерелом, часовим горизонтом і, за потреби, правилами інтерпретації. Інакше система швидко втратить масштабованість у разі розширення переліку платформ або появи нових метрик [14; 17; 18; 20; 22; 23; 24].

Для практичних потреб університету принципове значення має також багатоджерельний і комбінований характер даних, що підлягають обліку. Повноцінне уявлення про наукометричну активність неможливо сформулювати лише на основі однієї зовнішньої платформи. Необхідно поєднувати зовнішні відомості про публікації, цитування, профілі, індекси, DOI, ORCID та інституційні зв'язки з внутрішніми даними про підрозділи, посади, належність працівника до кафедри чи факультету, період роботи та інші організаційні характеристики. Відкриті інфраструктури на кшталт Crossref і OpenAlex надають багаті метадані про роботи, авторів та інституції, але саме внутрішні університетські системи додають управлінський контекст, без якого ця інформація не може бути повноцінно використана для звітності чи аналітики [25; 26; 28; 33].

Ще одна важлива риса — мережевий характер зв'язків усередині предметної області. Наукометрична активність не є простою сумою індивідуальних досягнень. Кожна публікація може мати кількох авторів, кожен автор — бути пов'язаним із кількома інституціями, одна робота — індексуватися в кількох базах, а її показники — змінюватися по-різному залежно від платформи. Крім того, існують тематичні, цитатні, організаційні та профільні зв'язки. OpenAlex спеціально моделює ці дані як граф взаємопов'язаних сутностей, а Scopus Author Profiles відображає також співавторства і пов'язані елементи наукового доробку. З погляду моделювання це означає, що предметна область потребує не плоскої

таблиці, а системи сутностей і зв'язків, серед яких значна частина має характер «багато до багатьох» [28; 32; 33].

Окремої уваги потребує питання якості та достовірності метаданих. Для наукометричного обліку важливо, щоб дані були не лише доступними, а й достатньо точними для подальшого використання в аналітиці та звітності. Crossref у своїх матеріалах прямо підкреслює значення якісних метаданих, зокрема авторських афіліацій та пов'язаних ідентифікаторів, а Elsevier наголошує на важливості актуального й точного авторського профілю. З цього випливає, що інформаційна система повинна не просто накопичувати дані, а й передбачати логіку їх перевірки, уточнення, оновлення та відокремлення підтверджених відомостей від неперевірених або автоматично імпортованих [32; 35].

Отже, наукометрична активність як об'єкт інформаційного обліку характеризується структурною складністю, різномірністю сутностей, багатоджерельністю, динамічністю показників, залежністю від платформи, потребою в надійній ідентифікації автора та високою чутливістю до якості метаданих. Саме це робить її специфічною предметною областю для спеціальності 126 «Інформаційні системи та технології». Для коректного обліку тут недостатньо простої фіксації набору числових значень або переліку публікацій. Потрібні чітко сформовані вимоги до системи, а також продумані моделі даних, здатні відобразити як стабільні, так і змінні характеристики наукової активності, зберегти контекст їх походження та створити підґрунтя для подальшої аналітики й розвитку системи.

1.4 Проблеми та перспективи автоматизації обліку наукометричної активності

Автоматизація обліку наукометричної активності в університетському середовищі сьогодні не зводиться до технічного збирання та відображення даних. Йдеться про створення інформаційного

середовища, здатного працювати з різнорідними джерелами, фіксувати змінні показники, забезпечувати ідентифікацію дослідників та підтримувати аналітичну придатність відомостей у тривалій перспективі. Проблематика автоматизації цієї сфери лежить на стику двох площин: коректного розуміння природи самих наукометричних даних та архітектурних рішень для їх системного опрацювання.

Ключові проблеми автоматизації обліку наукометричної активності виявлені в попередніх підрозділах, у цьому підрозділі вони синтезуються у трьох взаємопов'язаних напрямках. Перший — *гетерогенність джерел і метаданих*: інформація про діяльність дослідника формується в кількох незалежних системах (ORCID, Crossref, OpenAlex, Scopus, Web of Science), кожна з яких має власну логіку побудови, набір сутностей і формат метаданих. Інтеграція не може зводитися до копіювання окремих полів — потрібне узгодження різних структур, ідентифікаторів, рівнів деталізації та ступенів повноти [13; 26; 28]. Другий напрям — *неоднорідність і непорівнюваність показників*: близькі за назвою індикатори у Scopus, Google Scholar та OpenAlex можуть мати різні значення через відмінності в охопленні джерел і алгоритмах оновлення. Тому система повинна зберігати значення разом із указівкою на джерело, час фіксації, об'єкт застосування та контекст інтерпретації [17; 23; 28]. Третій напрям — *якість даних і відповідальне використання метрик*: автоматизація сама по собі не гарантує достовірності, а може масштабувати наявні в зовнішніх джерелах прогалини і дублікати. Лейденський маніфест і декларація DORA прямо застерігають проти спрощеного трактування метрик як універсального еквівалента наукової якості [15; 16; 35].

Окремо слід зазначити проблему динамічного характеру наукометричних даних. На відміну від багатьох статичних довідкових сутностей, значна частина показників наукової активності змінюється в часі: цитування зростають або коригуються, профілі доповнюються,

платформи оновлюють зв'язки між записами. Це означає, що автоматизована система повинна не просто зберігати один стан даних, а підтримувати часові зрізи, історію змін і повторну актуалізацію інформації — інакше вона швидко втрачає аналітичну цінність [26; 32; 34].

Перспективи автоматизації обліку наукометричної активності визначаються розвитком зрілих рішень класу CRIS (Current Research Information System) — інформаційних систем управління дослідницькою інформацією, які забезпечують облік публікацій, проєктів, профілів дослідників, афіліацій, фінансування та аналітики на рівні установи. У світовій практиці сформувалося три помітні моделі реалізації CRIS, що ілюструють різні підходи до архітектури та масштабованості [55; 56; 57]. Одним із найвідоміших комерційних рішень класу CRIS є Pure (Elsevier). Система забезпечує єдину точку доступу до даних про дослідницьку активність, підтримує профілі науковців, аналітичні дашборди та інтеграційні сценарії для установ, однак є закритим комерційним рішенням із відповідними обмеженнями щодо кастомізації [55]. Альтернативу складають відкриті рішення DSpace-CRIS (4Science) і VIVO. DSpace-CRIS має гнучку модель даних, повну сумісність з ORCID API v3, підтримує стандарти OpenAIRE Guidelines for CRIS Managers v1.1.1 на основі моделі CERIF, надає мережеві графи співавторства та агреговану бібліометрику [56]. VIVO використовує семантичний підхід на основі RDF, що робить його придатним для побудови зв'язаних графів дослідницької інформації, однак вимагає вищого рівня технічної підготовки команди супроводу [57].

Загальною перспективою для всіх типів CRIS є зростання ролі відкритих метаданих та API-інтеграції. REST API Crossref і відкриті дампи OpenAlex технічно роблять реалістичним автоматизоване отримання значної частини наукової інформації, що зменшує дублювання ручної роботи й підвищує оперативність актуалізації [26;

28]. Друга перспектива — підтримка узагальнених моделей даних, орієнтованих не на жорстко зафіксований перелік полів, а на набір взаємопов'язаних сутностей, типів зв'язків і атрибутів, що можуть розширюватися. Саме така архітектура застосована у DSpace-CRIS і відповідає логіці CERIF [56]. Третя перспектива — підтримка відповідального та контекстного аналізу: коли система зберігає не лише кінцеві значення метрик, а й джерело, дату фіксації, тип і рівень застосування показника, це створює підґрунтя для якіснішої аналітики, що відповідає сучасному баченню відповідальної наукометрії [15; 16].

Узагальнену порівняльну характеристику основних моделей CRIS і поточного стану *publication.kspu.edu* як вихідної точки розвитку подано в таблиці 1.2

Таблиця 1.2

Порівняльна характеристика основних моделей CRIS та
publication.kspu.edu

Характеристика	Pure (Elsevier)	DSpace-CRIS	VIVO	publication.kspu.edu
Тип ліцензії	комерційна	open-source	open-source	внутрішнє університетське рішення
Модель даних	пропрієтарна	CERIF, OpenAIRE	RDF, семантична	наявна, потребує формалізації та модернізації
Підтримка ORCID	повна	ORCID API v3	через розширення	часткова; потребує розширення інтеграції
Підтримка API	REST API	REST API	SPARQL	наявний JSON-обмін; потребує актуалізації та розширення
Інтеграція з Scopus / WoS	пряма	через імпорт	через імпорт	парсинг, API або змішані механізми залежно від джерела
Часові зрізи показників	підтримуються	підтримуються	підтримуються	частково представлені; потребують системної підтримки
Багаторівнева аналітика (кафедра, факультет, установа)	повна	повна	через шаблони	повна
Орієнтовний рівень складності впровадження	високий (комерційний контракт)	середній	високий (потребує фахівців з RDF)	середній: наявне рішення потребує модернізації

Порівняння показує, що *publication.kspu.edu* уже має суттєву функціональну й технологічну основу: рейтингові представлення, багаторівневу аналітику, роботу з наукометричними джерелами та

механізми обміну даними. Водночас, на відміну від зрілих CRIS-систем, подальший розвиток сервісу потребує формалізованого опису моделі даних, системної підтримки історичних зрізів, уніфікованого каталогу показників, прозорого фіксування джерел походження даних і модернізації інтеграційного шару. Це визначає основні напрями подальшого розвитку, які мають наблизити сервіс до архітектурних принципів сучасних CRIS, не перетворюючи його на копію Pure чи DSpace-CRIS — з огляду на специфіку університетського контексту й можливості підтримки.

Отже, проблеми автоматизації обліку наукометричної активності — гетерогенність джерел, непорівнюваність показників, динамічність даних і ризики спрощеного трактування метрик — формують конкретний запит на розвиток інформаційних систем класу CRIS. Аналіз існуючих рішень (Pure, DSpace-CRIS, VIVO) дозволяє визначити архітектурні орієнтири: підтримка стандартів CERIF або сумісних з ним моделей, повна інтеграція з ORCID, відкриті API, часові зрізи показників і відповідальне використання метрик. Саме ці орієнтири задають перспективу подальшого вдосконалення *publication.kspu.edu* і формують підставу для аналізу його практичного контексту функціонування у другому розділі роботи.

Висновки до розділу 1

У першому розділі досліджено теоретичні засади обліку наукометричної активності та особливості її автоматизації, що відповідає першому завданню роботи — дослідженню наукометрії як предметної області інформаційної системи. Розгляд її саме у такій

перспективі дозволив перейти від опису окремих показників і платформ до системного аналізу даних, джерел їх формування та принципів подальшого інформаційного моделювання, що відповідає профілю спеціальності 126 «Інформаційні системи та технології».

Аналіз сучасних наукометричних платформ показав, що цифрове середовище наукової комунікації формується сукупністю різнотипних інфраструктур, які відрізняються за функціональним призначенням, складом метаданих, доступністю API та набором підтримуваних показників. Розглянуто Scopus і Web of Science як комерційні реферативно-цитатні бази, Google Scholar як академічну пошукову платформу, ORCID як систему стійкої ідентифікації дослідників, Crossref і OpenAlex як відкриті інфраструктури метаданих, а також Dimensions як зв'язану аналітичну базу. З'ясовано, що ці системи надають дані різної природи й не можуть розглядатися як взаємозамінні джерела для університетської інформаційної системи.

Виявлено, що наукометричні показники є неоднорідними за змістом і рівнем застосування: одні характеризують автора (h-індекс, i10-індекс, кількість цитувань), інші — окрему публікацію (цитованість, альтметрики), треті — джерело публікації (CiteScore, SJR, SNIP, JIF) або інституційний рівень (агреговані рейтинги, мережеві характеристики співавторства). Однакові за назвою показники в різних платформах можуть мати різні значення через відмінності в охопленні джерел і алгоритмах оновлення. Обґрунтовано, що в інформаційній системі показник має зберігатися не як ізольоване числове поле, а як структурований об'єкт, пов'язаний з типом метрики, об'єктом застосування, платформою-джерелом, часовим зрізом і статусом верифікації.

Установлено, що наукометрична активність як об'єкт інформаційного обліку має складну багаторівневу природу. Вона охоплює авторські профілі, афіліації, зовнішні ідентифікатори (ORCID,

DOI, ROR, ISSN), структурні зв'язки, динамічні показники, часові зміни та мережеві взаємозв'язки між авторами, установами й джерелами. Для її коректного обліку недостатньо спрощеної моделі «автор — перелік публікацій» — потрібна система сутностей і зв'язків, здатна відобразити як стабільні, так і змінні характеристики предметної області, включно з підтримкою зв'язків типу «багато до багатьох» (співавторство) та поліморфних посилань (значення показника для різних типів об'єктів).

Виявлено основні проблеми автоматизації обліку наукометричної активності: гетерогенність джерел і метаданих, неповна порівнюваність показників між платформами, складність ідентифікації автора, динамічний характер метрик, ризики втрати якості метаданих та небезпека спрощеного трактування кількісних індикаторів. Принципи відповідального використання метрик, закріплені у Лейденському маніфесті та декларації DORA, прямо застерігають проти інтерпретації окремого показника як універсального еквівалента наукової якості, що має бути враховано в архітектурі будь-якої університетської системи рейтингування.

Проведений порівняльний аналіз існуючих рішень класу CRIS (Current Research Information System) — комерційного Pure (Elsevier), відкритих DSpace-CRIS (4Science) та VIVO — дозволив визначити архітектурні орієнтири для подальшого розвитку інформаційної системи обліку наукометричної активності. До них належать сумісність з моделлю даних CERIF та рекомендаціями OpenAIRE Guidelines for CRIS Managers, повна інтеграція з ORCID API v3, підтримка часових зрізів показників, відкриті REST-інтерфейси та механізми гнучкого розширення набору сутностей без структурної перебудови бази даних. Установлено, що сервіс publication.kspu.edu уже забезпечує важливу частину функціональності, релевантної для CRIS-підходу: багаторівневу аналітику, рейтингові представлення та роботу з наукометричними даними. Водночас його подальший розвиток потребує формалізованого

опису й модернізації моделі даних, системної підтримки часової фіксації показників та розширення API-орієнтованої інтеграції із зовнішніми джерелами.

Сформована в першому розділі теоретична та методична основа дозволяє перейти до аналізу практичного контексту функціонування `publication.ksru.edu` та формування функціональних і нефункціональних вимог до його подальшого розвитку, що становить зміст другого розділу роботи.

РОЗДІЛ 2 ФОРМУВАННЯ ВИМОГ ДО ІНФОРМАЦІЙНОЇ СИСТЕМИ ОБЛІКУ НАУКОМЕТРИЧНОЇ АКТИВНОСТІ

2.1 Аналіз практичного контексту та досвіду використання системи publication.kspu.edu

Перехід від загального аналізу наукометрії до формування вимог до інформаційної системи потребує розгляду реального практичного контексту, у якому така система функціонує. У межах цієї дипломної роботи таким контекстом є електронний сервіс publication.kspu.edu, що використовується в Херсонському державному університеті для подання, накопичення та аналітичного відображення відомостей про наукові публікації й окремі показники публікаційної активності. На офіційних сторінках університету цей ресурс прямо визначається як електронний сервіс рейтингування, а його впровадження пов'язується з наказом Херсонського державного університету від 23.04.2020 № 385-Д. Там само зазначено, що сервіс призначений для рейтингування за адресою publication.kspu.edu, а автором ідеї та методології вказано проф. О. Співаковського [3; 36; 37].

Аналіз публічно доступних розділів системи свідчить, що вона орієнтована не лише на збереження окремих записів про публікації, а й на аналітичне представлення наукової активності на різних рівнях. У відкритій частині ресурсу наявні розділи «Аналітика», «Рейтинг факультетів», «Рейтинг кафедр», «Рейтинг науково-педагогічних працівників», відомості за галузями знань, а також інформація про заклади-партнери. Це дає підстави стверджувати, що система вже функціонує не як проста бібліографічна база, а як інструмент структурованого представлення наукової активності університету та пов'язаних із ним суб'єктів [3; 5; 6; 36; 38; 39; 40].

Важливо й те, що система реалізує багаторівневий аналітичний підхід. Із відкритих сторінок видно, що в сервісі відображаються

результати не лише на рівні установи в цілому, а й на рівні факультетів, кафедр та окремих науковців. Сторінки рейтингування містять агреговані представлення показників за підрозділами, а персональні сторінки окремих осіб можуть включати відомості, пов'язані зі Scopus, Google Scholar і Web of Science. Для цієї дипломної роботи це має принципове значення, оскільки показує: навіть на поточному етапі функціонування publication.kspu.edu вже містить елементи інтеграції кількох платформ і підтримує не лише документний, а й профільно-аналітичний рівень подання даних [6; 17; 20; 23; 32; 36; 38; 39].

Важливо врахувати, що publication.kspu.edu має попередньо описану технологічну історію розвитку. У публікаціях, присвячених системі автоматичного рейтингування на основі наукометричних показників, описано архітектуру, що включала отримання даних із зовнішніх наукометричних систем, системну базу даних, SQL-обробку, API-доступ, JSON-обмін та окремі аналітичні модулі. Тому в межах цієї роботи publication.kspu.edu розглядається не як система без інформаційної основи, а як наявне університетське рішення, що потребує подальшої формалізації, модернізації моделі даних, розширення історичної аналітики, верифікації та інтеграційного шару.

Окремої уваги заслуговує те, що основна практична цінність системи полягає у формуванні єдиного університетського простору представлення наукової активності. У новинному матеріалі про сервіс зазначено, що сайт дає змогу переглядати показники публікаційної активності як окремих учених, так і цілих наукових установ України, а також дізнаватися показники цитованості їхніх робіт. Це означає, що publication.kspu.edu від початку замислювався не лише як локальний реєстр публікацій, а як цифрове середовище, у якому наукова активність може бути агрегована, зіставлена й представлена в аналітично придатному вигляді [4; 36].

Для спеціальності 126 «Інформаційні системи та технології» особливо важливо, що ця система демонструє вже сформовану предметну область і водночас виявляє межі, в яких постає потреба в її подальшому розвитку. З одного боку, наявність аналітичних сторінок, рейтингових представлень, персональних профілів і даних за галузями знань свідчить про практичну спроможність системи працювати з багатьма сутностями та кількома рівнями узагальнення. З іншого боку, саме винесення в окремі розділи різних типів аналітики та різних платформ указує на потребу в такій моделі даних, яка не буде жорстко прив'язана до одного формату представлення або до обмеженого набору показників. Це підводить до висновку, що подальший розвиток системи має спиратися на узагальнення критеріїв представлення даних і на більш системний підхід до вимог та моделювання [3; 5; 6; 36; 38; 39; 40].

Ще однією важливою обставиною є те, що publication.kspu.edu не існує ізольовано від загальної цифрової інфраструктури університету. На офіційному сайті Херсонського державного університету сервіс прямо пов'язується з науковою та публікаційною діяльністю, а в університетських матеріалах простежується загальна логіка цифрового представлення публікацій, наукових робіт, аналітики та пов'язаних ресурсів. Це дає підстави розглядати систему не як випадковий окремий вебсайт, а як елемент університетського інформаційного середовища, орієнтованого на підтримку наукової діяльності. У контексті дипломної роботи це означає, що вимоги до системи доцільно формувати з урахуванням її ролі в ширшій цифровій архітектурі закладу вищої освіти [36; 37; 41; 42].

Разом із тим практичний досвід використання publication.kspu.edu виявляє й низку обмежень, важливих для подальшого проектування. Персональні сторінки вже можуть включати дані з кількох джерел, зокрема зі Scopus, Google Scholar і Web of Science, тобто система функціонує в середовищі, де показники надходять із різних платформ і

мають різну природу. За таких умов недостатньо просто додавати нові поля або окремі інформаційні блоки до наявного інтерфейсу. Необхідним стає перегляд підходів до структурування сутностей, зв'язків і самих показників. Інакше кожне розширення кількості платформ або метрик ускладнюватиме супровід системи та зменшуватиме її гнучкість.

Отже, аналіз практичного контексту використання `publication.kspu.edu` дає підстави для кількох висновків. По-перше, в університеті вже існує реальний досвід цифрового обліку та представлення наукової активності, що має прикладну цінність і підтверджує актуальність обраної теми. По-друге, система демонструє наявність кількох рівнів інформаційного подання — від персонального профілю до аналітики на рівні підрозділів і установ. По-третє, структура ресурсу та його функціональні елементи вказують на доцільність подальшого розвитку системи в напрямі розширення платформ, узгодження показників і вдосконалення моделей даних. Саме тому `publication.kspu.edu` у межах цієї роботи доцільно розглядати не як завершене рішення, а як практичну основу для формування вимог до сучаснішої інформаційної системи обліку наукометричної активності.

2.2 Обґрунтування напрямів розвитку інформаційної системи обліку наукометричної активності

Після аналізу предметної області та практичного контексту використання `publication.kspu.edu` закономірно постає питання не лише про вдосконалення окремих елементів сервісу, а про визначення загальних напрямів його подальшого розвитку. Таке обґрунтування повинно спиратися на дві групи чинників: реальні потреби університетського середовища в обліку, аналітиці та представленні наукової активності, а також сучасні можливості зовнішніх наукометричних платформ, відкритих метаданих і цифрових

ідентифікаторів. Саме поєднання цих чинників формує підстави для переходу від наявного сервісу до більш гнучкої інформаційної системи, орієнтованої на розширення даних, уніфікацію показників і підвищення аналітичної цінності результатів [13; 26; 28; 36; 41]. Для уточнення відмінності між наявною інформаційною основою publication.kspu.edu та запропонованою в роботі моделлю даних доцільно подати їх порівняльну характеристику за ключовими критеріями розвитку системи (табл. 2.1).

Таблиця 2.1

Порівняння наявної інформаційної основи publication.kspu.edu та запропонованої моделі даних

Критерій	Наявна реалізація publication.kspu.edu	Запропонована модель у роботі
Загальна роль бази даних	Використовується як інформаційна основа для збереження й обробки даних сервісу рейтингування.	Розглядається як формалізована модель подальшого розвитку системи з чітко визначеними сутностями, зв'язками, ключами та обмеженнями цілісності.
Структура даних	Забезпечує роботу наявних функцій сервісу: рейтинги, персональні сторінки, аналітичні представлення.	Передбачає розширену структуру сутностей: науковці, профілі, публікації, платформи, показники, значення показників, джерела даних, часові зрізи.
Представлення показників	Орієнтоване на обробку й відображення наявних наукометричних показників.	Показники подаються через універсальну модель <code>MetricTypes</code> і <code>MetricValues</code> , що дозволяє додавати нові метрики без перебудови центральної структури БД.
JSON та обмін даними	JSON використовується як формат обміну даними та взаємодії з іншими компонентами або зовнішніми системами.	JSON розглядається як формат API-обміну, а отримані JSON-метадані перетворюються у структуровану реляційну модель.
Інтеграція з джерелами	Підтримується через наявні механізми отримання даних, зокрема парсинг, API або змішаний підхід залежно від джерела.	Передбачається розвиток уніфікованого інтеграційного шару для роботи з Crossref, OpenAlex, ORCID та іншими джерелами.
Історичні зрізи	Частково можуть бути представлені в аналітичній	Виділяються як окремий елемент моделі через Snapshots /

Критерій	Наявна реалізація publication.kspu.edu	Запропонована модель у роботі
	логіці системи.	TimeSlices для системного збереження динаміки показників.
Верифікація даних	Може здійснюватися в межах поточної логіки роботи системи, але потребує чіткішої формалізації.	Передбачено фіксацію статусу верифікації, джерела походження та способу отримання запису.
Масштабування	Розширення можливе, але може потребувати додаткових змін у логіці обробки даних.	Розширення платформ і метрик виконується через довідники Platforms, DataSources, MetricTypes без руйнування центральної моделі.
Основне призначення в контексті роботи	Практична основа, на якій уже функціонує університетський сервіс рейтингування.	Проектна модель модернізації, яка уточнює структуру даних і створює підґрунтя для подальшої реалізації оновленої системи.

Отже, запропонована в роботі модель не заперечує наявності бази даних у попередній реалізації publication.kspu.edu. Навпаки, наявна система розглядається як практична інформаційна основа, що вже забезпечує збереження, обробку та аналітичне представлення наукометричних даних. Водночас подальший розвиток сервісу потребує формалізованого опису моделі даних, уніфікованого представлення показників, системної підтримки часових зрізів, фіксації джерел походження та статусів верифікації даних. Саме ці завдання і становлять зміст запропонованої в роботі моделі.

Першим напрямом розвитку є розширення спектра джерел наукометричних даних. Уже сам практичний досвід publication.kspu.edu показує, що система функціонує в логіці багаторівневого аналітичного представлення наукової активності, а отже її подальший розвиток неминуче пов'язаний із потребою працювати не лише з обмеженим набором платформ. Сучасне наукове середовище пропонує джерела, які відрізняються не тільки набором показників, а й типами доступних метаданих: ORCID забезпечує стійку ідентифікацію автора, Crossref надає відкриті бібліографічні метадані через REST API, а OpenAlex

функціонує як відкритий каталог взаємопов'язаних сутностей глобальної дослідницької системи. Це означає, що перспективна інформаційна система обліку наукометричної активності повинна проектуватися як така, що принципово здатна інтегрувати кілька джерел, а не як система, прив'язана до фіксованого набору платформ [13; 26; 28; 36].

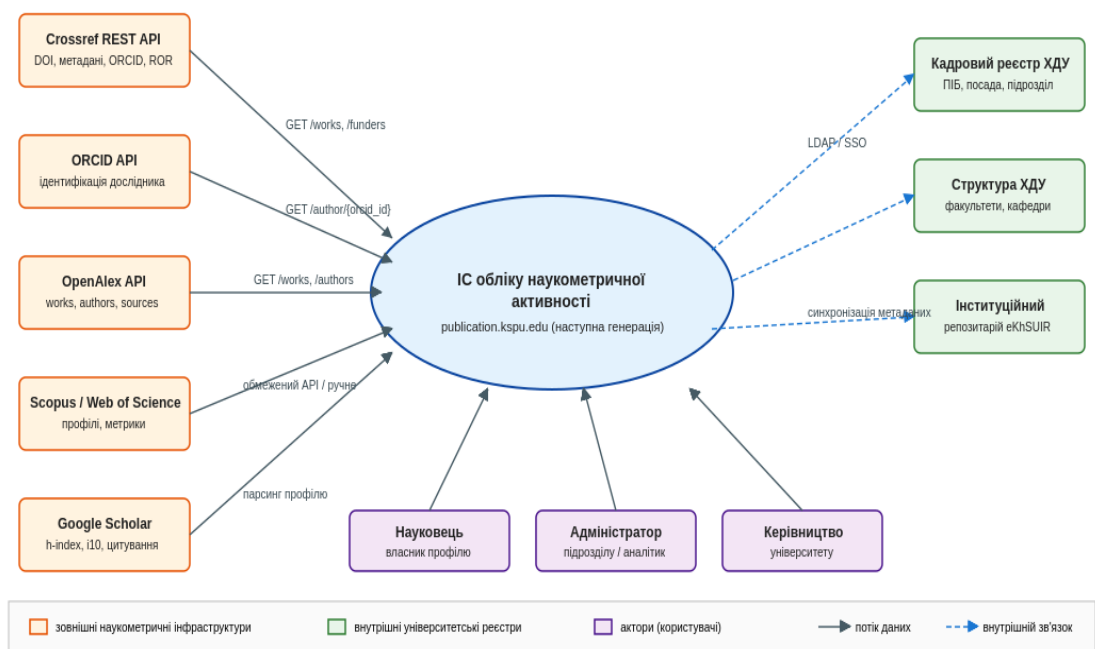
Другий напрям розвитку пов'язаний із переходом від зберігання окремих показників до узагальненого представлення критеріїв наукометричної оцінки. У сучасних платформах одна й та сама наукова активність описується різними метриками: частина з них стосується автора, частина — публікації, частина — джерела публікації або інституції. Крім того, навіть подібні показники можуть мати різний зміст залежно від джерела. Тому розвиток системи доцільно спрямувати не на механічне додавання нових полів під кожен метрику, а на створення такої моделі даних, у якій показник описується через спільні атрибути: назву, тип, джерело, рівень застосування, часовий зріз, одиницю вимірювання та значення. Саме такий підхід забезпечує масштабованість системи й дозволяє розширювати перелік платформ без постійного структурного перепроєктування [17; 20; 23; 24; 26; 28; 29].

Окремого значення набуває посилення ролі цифрової ідентифікації дослідника. У практиці обліку наукової активності одна з найстійкіших проблем пов'язана з ототожненням автора в різних системах, особливо за наявності варіантів написання імені, змін афіліації або дублювання профілів. ORCID прямо визначає свою місію як забезпечення прозорих і надійних зв'язків між дослідниками, їхнім внеском та афіліаціями завдяки унікальному постійному ідентифікатору. Тому одним із ключових напрямів розвитку інформаційної системи має стати побудова моделі, у якій ідентифікатори автора не є другорядним доповненням, а виступають центральним механізмом зв'язування внутрішніх і зовнішніх даних. Для університетського середовища це дає змогу зменшити кількість дублювань, підвищити точність зіставлення профілів

і поліпшити якість аналітики на рівні окремих науковців та підрозділів [13; 25].

Наступний напрям розвитку полягає в орієнтації системи на роботу з відкритими метаданими та API-інтеграцією. Сучасні цифрові наукові інфраструктури дедалі більше підтримують машинно-читане отримання даних. Зокрема, Crossref через REST API надає не лише базові бібліографічні метадані, а й інформацію про фінансування, ліцензії, ORCID, ROR, анотації та інші пов'язані атрибути. Це відкриває нові можливості для розвитку університетських інформаційних систем: від ручного або напівручного введення можна перейти до інтегрованого отримання й оновлення частини даних. У практичному вимірі це означає підвищення оперативності актуалізації відомостей, зменшення дублювання ручної роботи та формування більш сучасної архітектури системи, орієнтованої на повторне використання зовнішніх відкритих даних. Узагальнену логіку взаємодії інформаційної системи із зовнішніми наукометричними інфраструктурами та внутрішніми університетськими реєстрами доцільно подати у вигляді контекстної діаграми (рис. 2.1) [26; 27].

Рисунок 2.1 — Контекстна діаграма взаємодії
інформаційної системи обліку наукометричної активності



Важливим напрямом є і розвиток аналітичної складової системи. Уже наявна структура publication.kspu.edu свідчить, що сервіс орієнтується не лише на збереження відомостей, а й на аналітичне подання результатів — через рейтинги факультетів, кафедр, науково-педагогічних працівників, галузей знань і партнерських установ. Це означає, що подальший розвиток системи логічно пов'язаний не стільки з накопиченням дедалі більшої кількості записів, скільки з підвищенням якості аналітичного представлення даних. У перспективі це може передбачати підтримку часових зрізів, ширших порівняльних вибірок, прозорішого зв'язку між первинними даними та агрегованими результатами, а також можливості пояснюваного аналізу показників [3; 5; 6; 36; 38; 39; 40].

Окремим напрямом розвитку є поєднання зовнішніх наукометричних даних із внутрішнім університетським контекстом. Зовнішні платформи дають змогу бачити публікації, профілі, цитування та індекси, однак самі по собі вони не враховують повною мірою організаційної структури закладу вищої освіти, кадрового контексту, належності працівника до певної кафедри чи факультету, часових меж його роботи в установі. Саме внутрішній контекст є критичним для управлінської аналітики, підготовки звітності та внутрішнього моніторингу. Тому система має розвиватися як інструмент, що поєднує зовнішню наукометричну інформацію з внутрішніми структурними даними університету. У цьому полягає одна з її ключових відмінностей від зовнішніх платформ загального призначення [36; 41; 42].

Ще одним необхідним напрямом розвитку є підвищення достовірності та керованості даних. Автоматизація не усуває проблему якості інформації; навпаки, вона робить її більш помітною. Якщо система працює з кількома джерелами, потрібно чітко розрізняти автоматично отримані, підтверджені, уточнені та потенційно неповні записи. У відкритих наукових інфраструктурах багатство метаданих

поєднується з їх різною повнотою, а отже модель майбутньої системи повинна враховувати не лише самі сутності й атрибути, а й походження даних, час їх отримання, статус верифікації та ступінь довіри. У прикладному сенсі це важливо для того, щоб аналітичні результати ґрунтувалися на контрольованій інформаційній основі, а не на непрозорому наборі агрегованих значень [26; 27; 35].

Окремим перспективним напрямом подальшого розвитку системи може стати залучення засобів штучного інтелекту для розв'язання допоміжних завдань, пов'язаних із верифікацією профілів науковців, нормалізацією метаданих, тематичною класифікацією публікацій та розширенням аналітичних можливостей. Водночас реалізація таких рішень потребує модернізації серверної частини, формалізації та розширення моделі бази даних, системного накопичення історичних зрізів і забезпечення контрольованого структурованого збереження даних. Саме це формує технологічну основу для подальших етапів розвитку системи [26; 28; 32; 34].

Отже, обґрунтовані напрями розвитку інформаційної системи обліку наукометричної активності пов'язані не з локальним розширенням окремих функцій, а з переходом до більш цілісної моделі системи. Її подальший розвиток має ґрунтуватися на інтеграції кількох джерел даних, уніфікованому представленні показників, посиленні ролі цифрових ідентифікаторів, використанні відкритих метаданих, розвитку аналітичної складової, поєднанні зовнішніх і внутрішніх даних та підвищенні керованості інформації. Саме така логіка створює підстави для наступного етапу дослідження — формування конкретних функціональних вимог до системи.

2.3 Формування функціональних вимог до системи

Сформовані в попередніх підрозділах загальні уявлення про предметну область, практичний досвід використання *publication.kspu.edu*

та обґрунтовані напрями розвитку доцільно перевести у форму конкретних можливостей системи. Під функціональними вимогами в цій роботі розуміється сукупність дій, операцій і сервісів, які інформаційна система обліку наукометричної активності повинна забезпечувати для підтримки накопичення, узгодження, аналізу та представлення даних. Із практичного боку формування таких вимог спирається на досвід функціонування *publication.kspu.edu* як сервісу рейтингування Херсонського державного університету, а з технологічного — на можливості сучасних відкритих наукових інфраструктур, насамперед ORCID, Crossref та OpenAlex, які надають структуровані дані, придатні для інтеграції в університетські цифрові рішення [13; 26; 28; 36].

Виявлені вимоги доцільно групувати у чотири логічні блоки залежно від аспекту функціонування системи: облік предметної області, інтеграція із зовнішніми джерелами, аналітика й часовий контекст, керованість і доступ. Такий поділ полегшує подальше проектування й дозволяє уникнути дублювання вимог одна на одну.

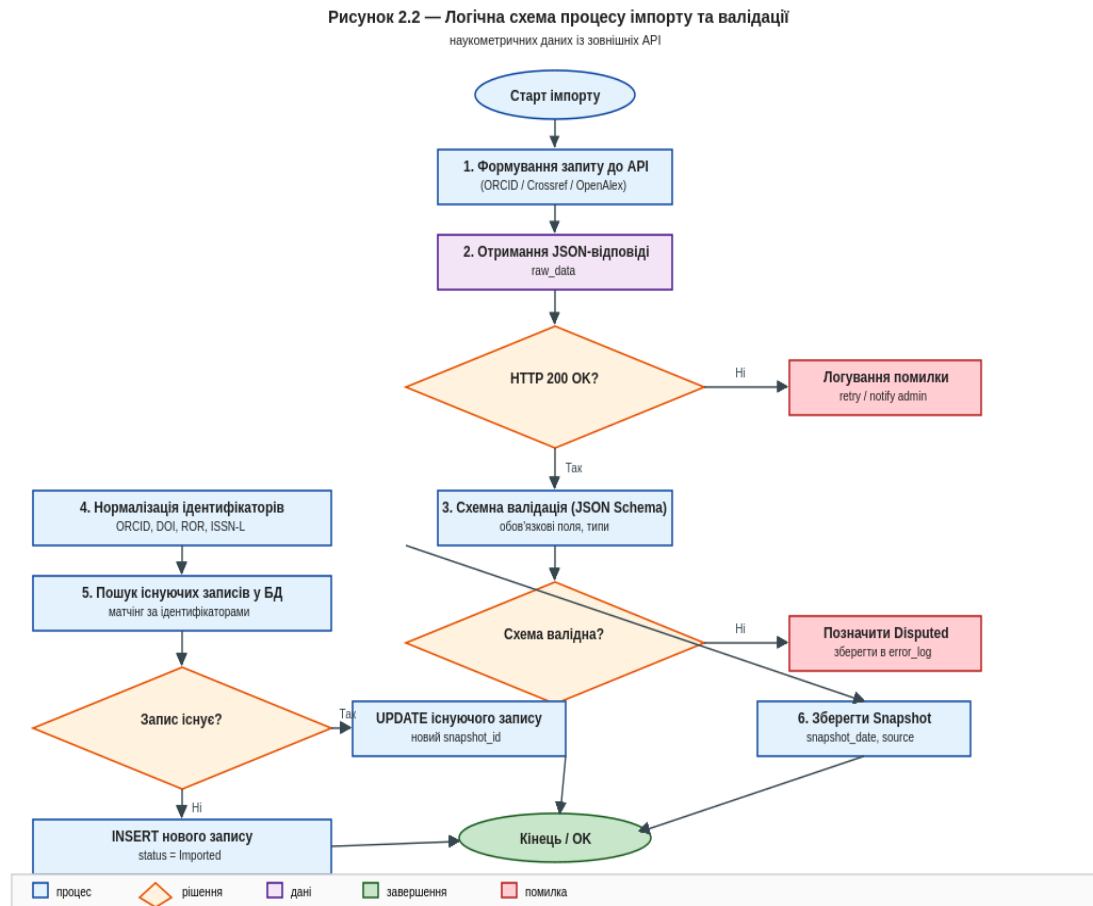
Перший блок охоплює вимоги, пов'язані з обліком базових сутностей предметної області (FR-01, FR-02, FR-03). Система має зберігати й опрацьовувати дані про публікації, науковців, авторські профілі, афіліації, структурні підрозділи, наукометричні платформи та показники, що впливає з характеру самої предметної області, описаної в підрозділі 1.3. Центральним інформаційним об'єктом виступає профіль науковця, який поєднує персональні дані, афіліації, зовнішні ідентифікатори, перелік наукових результатів і пов'язані метрики. Така побудова відповідає логіці ORCID, де дослідник представлений через постійний унікальний ідентифікатор, пов'язаний із його внеском у науку, а також узгоджується з підходом *publication.kspu.edu*, де персональні сторінки та рейтингування науково-педагогічних працівників уже є наявними елементами сервісу [6; 13; 36]. Окремою важливою вимогою є повноцінна робота з публікаціями: бібліографічний опис, цифрові

ідентифікатори, тип документа, авторський склад, джерело оприлюднення та інші метадані повинні зберігатися як пов'язані атрибути, а не як суцільне текстове поле [26].

Другий блок становлять вимоги до інтеграції із зовнішніми джерелами наукометричних даних (FR-04, FR-08). Сучасна архітектура наукометричних сервісів передбачає машинно-читане отримання інформації, тому університетська система не повинна обмежуватися ручним введенням. Crossref надає REST API з бібліографічними метаданими, серед яких можуть бути ORCID, ROR, відомості про фінансування та ліцензії; OpenAlex функціонує як відкритий каталог взаємопов'язаних сутностей глобальної дослідницької системи [26; 28]. Звідси випливає вимога підтримки автоматизованого або напівавтоматизованого імпорту, зіставлення та оновлення зовнішніх записів. Невід'ємною частиною цієї логіки є робота зі стійкими ідентифікаторами — ORCID, DOI, ROR, ISSN — як механізмами зв'язування внутрішніх записів із зовнішніми сутностями. Без надійної ідентифікації автора облік буде вразливим до дублювання й помилок зіставлення, особливо у випадках варіативності написання прізвища, зміни афіліації або наявності кількох профілів однієї особи [13; 35].

Третій блок об'єднує вимоги до аналітичної функціональності та часового контексту даних (FR-05, FR-06, FR-07). Система має зберігати показники не просто як числові значення, а як інформаційні об'єкти з атрибутами типу метрики, об'єкта застосування, платформи-джерела, часового зрізу та одиниці вимірювання — інакше додавання нової метрики потребуватиме структурної зміни бази даних. Поряд із цим необхідна підтримка багаторівневої аналітики: на рівні окремого науковця, кафедри, факультету, галузі знань та установи в цілому, що відповідає реальній структурі публічних розділів *publication.kspu.edu* — рейтингів факультетів і кафедр, аналітики, розрізу за галузями знань [3; 5; 6; 38; 39; 40]. Третя вимога цього блоку — часова фіксація даних:

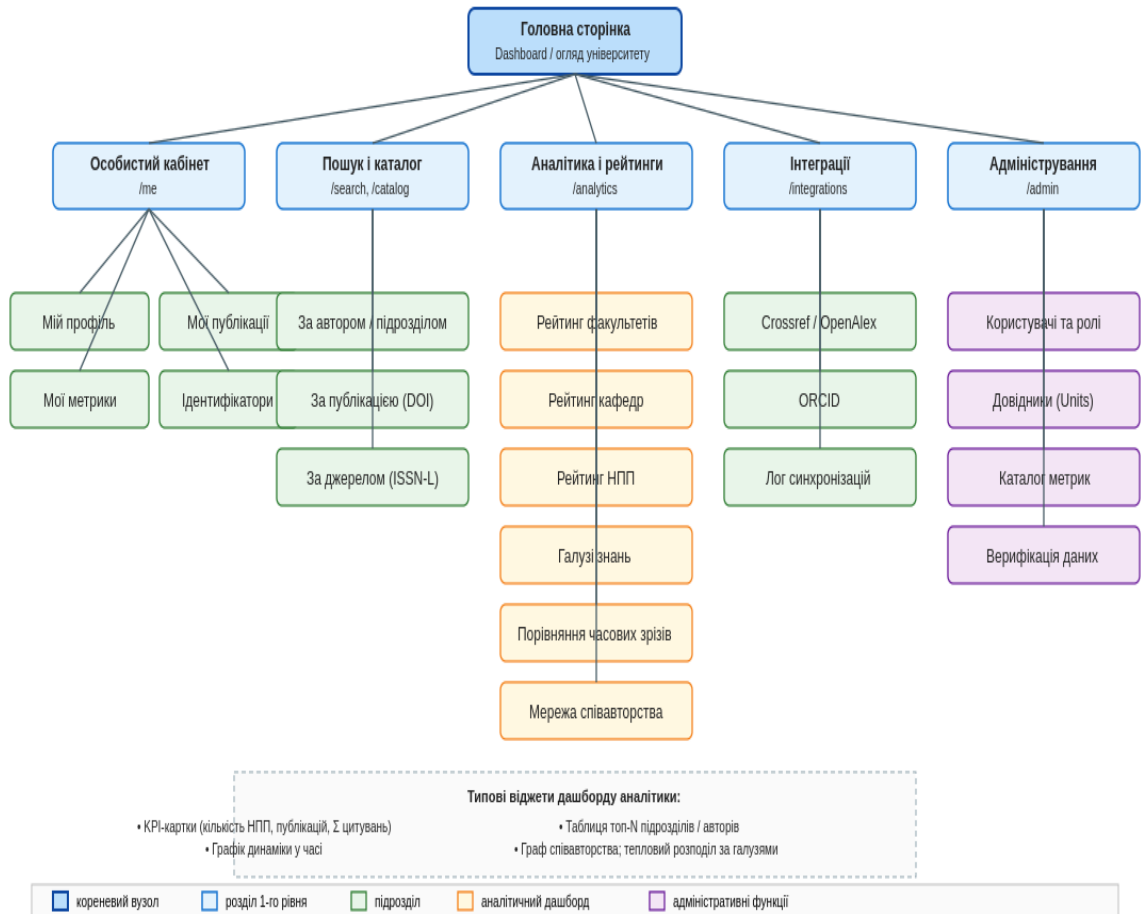
цитування, агреговані профілі та інші динамічні характеристики змінюються в зовнішніх системах, тому без збереження дати отримання значення та, за потреби, історії змін аналітика обмежиться випадковим «знімком» даних і не дозволить досліджувати тенденції [32; 34]. Базовий процес імпорту та валідації наукометричних даних із зовнішніх API подано на рисунку 2.2 у вигляді логічної схеми.



Четвертий блок становлять вимоги до керованості системи: верифікація даних, пошук і фільтрація, розмежування ролей користувачів (FR-08, FR-09, FR-10, FR-11). Записи можуть надходити з різних джерел і мати різний ступінь повноти, тому потрібні механізми перегляду, уточнення, підтвердження та коригування. Автоматичне отримання не скасовує контролю якості — особливо для університетської системи, де результати обліку використовуються для внутрішнього аналізу, рейтингування та звітності [35]. Пошук і фільтрація мають охоплювати авторів, підрозділи, типи публікацій,

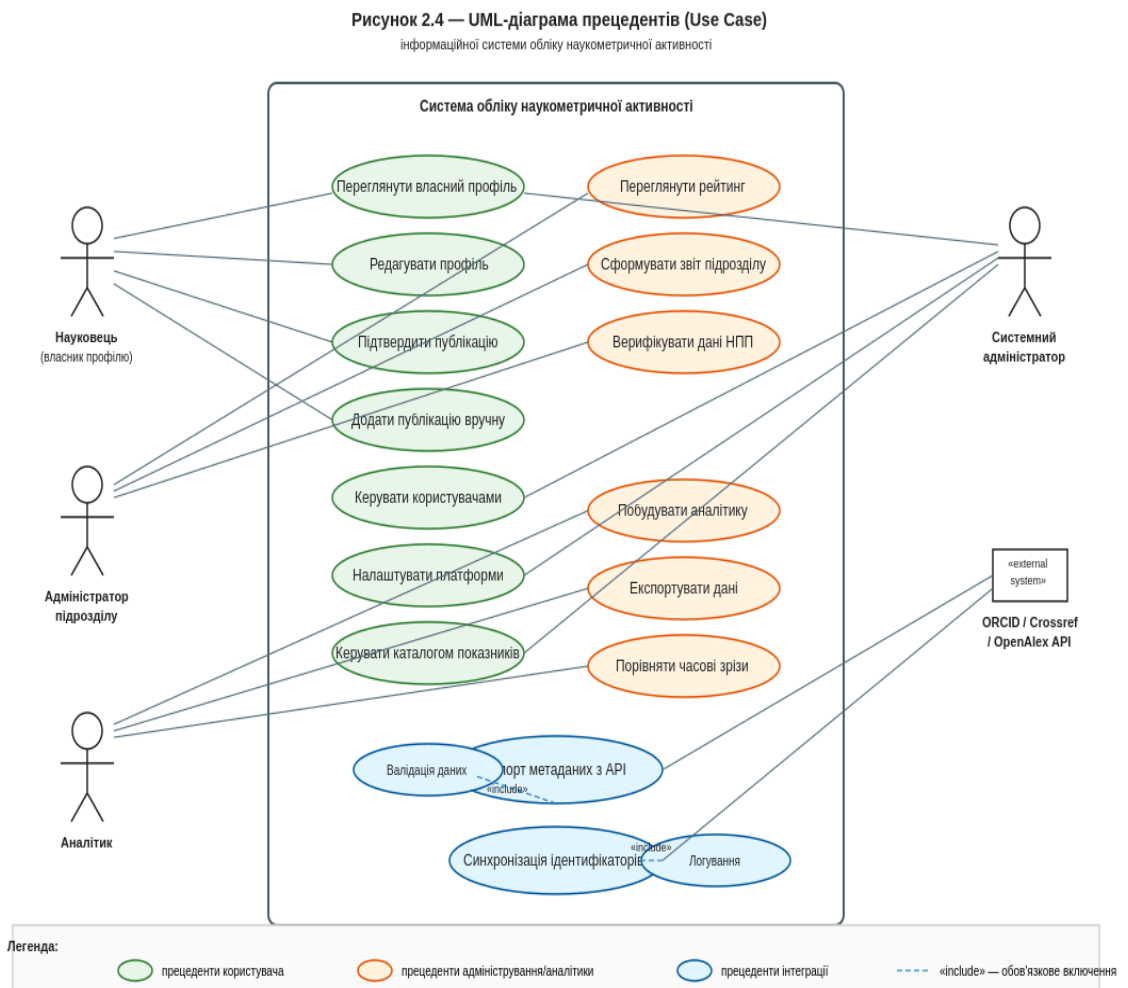
платформи, показники, галузі знань і часові періоди; саме така модель пошуку реалізована у відкритих API Crossref та OpenAlex і доцільна для майбутньої системи [26; 28]. Структуру навігації та аналітичних дашбордів розроблюваної системи показано на рисунку 2.3.

Рисунок 2.3 — Структура навігації та аналітичних дашбордів
інформаційної системи обліку наукометричної активності



Розмежування ролей користувачів зумовлене тим, що система працює одночасно з персональними, структурними та аналітичними даними. У цьому контексті виокремлюються базові ролі: науковець (перегляд та уточнення власного профілю), адміністратор підрозділу або аналітик (робота з агрегованими даними), системний адміністратор (керування довідниками, верифікація імпортованих записів), а також роль зовнішнього API як окремого «користувача» інтеграційних сценаріїв. Хоча публічна частина *publication.kspu.edu* відображає переважно відкриту аналітику, логіка функціонування сервісу рейтингування передбачає й керовані операції, тому вимога

розмежування доступу є обґрунтованою на рівні архітектури [36; 37]. Узагальнений розподіл функціональних можливостей між категоріями користувачів та зовнішніми модулями подано на UML-діаграмі прецедентів (рис. 2.4).



Зведену специфікацію сформованих функціональних вимог до інформаційної системи подано в таблиці 2.2.

Таблиця 2.2

Специфікація базових функціональних вимог до інформаційної системи

Ідентифікатор	Назва вимоги	Короткий опис функціональності
FR-01	Облік базових сутностей	Збереження даних про науковців, публікації, платформи та підрозділи
FR-02	Ведення профілю науковця	Управління персональними даними, афіліаціями та ідентифікаторами особи
FR-03	Управління публікаціями	Реєстрація бібліографічних описів робіт та авторського складу

Ідентифікатор	Назва вимоги	Короткий опис функціональності
FR-04	API-інтеграція	Автоматизоване отримання метаданих із зовнішніх відкритих джерел
FR-05	Облік метрик	Збереження різних типів показників із прив'язкою до об'єктів та платформ
FR-06	Багаторівнева аналітика	Формування зведених даних на рівні автора, кафедри, факультету, установи
FR-07	Часова фіксація	Підтримка часових зрізів для відстеження динаміки показників
FR-08	Верифікація даних	Механізми підтвердження, уточнення та коригування імпортованих записів
FR-09	Пошук та фільтрація	Навігація за даними, платформами, галузями знань та роками
FR-10	Формування рейтингів	Побудова підсумкових представлень і порівняльних таблиць
FR-11	Керування доступом	Розмежування прав доступу між ролями користувачів

Для реалізації функціональної вимоги FR-11 розроблено матрицю прав доступу (CRUD-модель), що регламентує повноваження базових ролей щодо ключових сутностей системи (Додаток А). Сформований перелік функціональних вимог охоплює базові операції з даними, інтеграцію із зовнішніми джерелами, роботу з профілями та публікаціями, підтримку часової динаміки, верифікацію, аналітичне представлення, пошук і керування доступом, що створює функціональну основу для подальшого визначення нефункціональних характеристик системи.

2.4 Формування нефункціональних вимог до системи

Поряд із функціональними характеристиками інформаційна система обліку наукометричної активності потребує специфікації нефункціональних вимог, які визначають якість її роботи, умови експлуатації, надійність, захищеність та придатність до подальшого розвитку. Йдеться не про ізольований програмний модуль, а про

систему, яка функціонує в університетському цифровому середовищі, взаємодіє із зовнішніми наукометричними сервісами та підтримує аналітичне використання результатів. Структурною основою для специфікації нефункціональних вимог у цій роботі обрано модель якості програмних продуктів ISO/IEC 25010, яка визначає вісім ключових характеристик: функціональну придатність, продуктивність, сумісність, зручність використання, надійність, захищеність, супроводжуваність і переносимість [43; 44]. Для специфікації якості даних додатково використано модель ISO/IEC 25012, що визначає характеристики структурованих даних у інформаційних системах [50]. Вимоги в галузі захищеності спираються на рекомендації OWASP щодо автентифікації, авторизації, керування сесіями та безпеки REST API [45; 46; 48; 49].

З огляду на специфіку предметної області з характеристик ISO/IEC 25010 для системи обліку наукометричної активності найбільш значущими є продуктивність, надійність, захищеність, сумісність, супроводжуваність і зручність використання; до них доданий блок якості даних на основі ISO/IEC 25012 та окрема вимога масштабованості, що впливає з потреби розширення переліку платформ і метрик. Сформовані нефункціональні вимоги наведено в таблиці 2.3 з ідентифікаторами NFR-01...NFR-09 і прив'язкою до відповідних характеристик стандартів.

Таблиця 2.3

Специфікація нефункціональних вимог до системи

Ідентифікатор	Категорія за ISO/IEC 25010	Назва вимоги	Цільовий показник
NFR-01	Performance efficiency	Час відгуку інтерфейсу	Не більше 2 с для перегляду профілю та рейтингу при ≤ 100 одночасних користувачах
NFR-02	Performance efficiency	Час виконання аналітичних запитів	Не більше 5 с для агрегованих

Ідентифікатор	Категорія за ISO/IEC 25010	Назва вимоги	Цільовий показник
			рейтингів на рівні факультету
NFR-03	Reliability	Цілісність даних	Збереження референтної цілісності та узгодженості при імпорті з зовнішніх API
NFR-04	Security (OWASP)	Автентифікація і авторизація	Розмежування 4 ролей (науковець, адмін підрозділу, системний адмін, зовнішній API), захищені сесії, OAuth 2.0 для API-інтеграції
NFR-05	Compatibility	Сумісність із зовнішніми сервісами	Підтримка ORCID, DOI, ROR, ISSN; обмін даними у форматі JSON через REST
NFR-06	Maintainability	Супроводжуваність моделі даних	Розширення переліку платформ і метрик без зміни структури центральних таблиць
NFR-07	Usability	Зручність використання	Підтримка трьох сценаріїв: персональний (науковець), управлінський (адмін підрозділу), системний (адміністратор)
NFR-08	Data Quality (25012)	Якість даних	Контроль повноти, точності, узгодженості та актуальності записів; фіксація

Ідентифікатор	Категорія за ISO/IEC 25010	Назва вимоги	Цільовий показник
			статусу верифікації
NFR-09	Scalability	Масштабованість	Підтримка зростання обсягу записів до 10^6 та розширення набору сутностей без структурної перебудови БД

Розглянемо обґрунтування ключових вимог із таблиці. Вимоги до продуктивності (NFR-01, NFR-02) впливають із того, що сервіс publication.ksru.edu вже містить аналітичні сторінки та рейтингові розділи, тобто користувач працює не лише з окремими записами, а й з агрегованими вибірками. Зовнішні джерела на кшталт Crossref REST API підтримують пошук, фільтрацію та вибірки, а архітектура інтеграції передбачає значну кількість запитів і обробку метаданих [26]. Цільові часи відгуку (2 секунди для типових операцій, 5 секунд для агрегатів) встановлені на основі загальноприйнятих орієнтирів для вебзастосунків і дозволяють запобігти неприйнятному уповільненню при зростанні обсягу записів та кількості платформ.

Вимога надійності (NFR-03) має особливу вагу, оскільки дані системи можуть використовуватися для внутрішнього аналізу, звітності та рейтингування. У прикладному вимірі надійність означає стійкість до помилок під час імпорту зовнішніх метаданих, коректну обробку неповних або змінених записів та збереження узгодженості між профілями, публікаціями, показниками й структурними одиницями. Це досягається через систему первинних і зовнішніх ключів, обмеження унікальності та правила референтної цілісності, описані в логічній моделі (підрозділ 3.3).

Вимога захищеності (NFR-04) спирається на рекомендації OWASP Cheat Sheet Series щодо автентифікації, авторизації, керування сесіями та

безпеки REST API [45; 46; 48; 49]. Система містить облікові записи користувачів, адміністративні функції та механізми редагування, тому потребує чіткого розмежування ролей (визначених у матриці CRUD у Додатку А), захисту адміністративних дій і безпечної взаємодії із зовнішніми API. Для інтеграційних сценаріїв з ORCID та іншими сервісами, що підтримують OAuth-протоколи, передбачено застосування рекомендацій OWASP щодо OAuth 2.0 [47].

Вимога якості даних (NFR-08) специфікована за моделлю ISO/IEC 25012, яка описує характеристики структурованих даних у інформаційних системах [50]. Для системи обліку наукометричної активності якість даних не є другорядною характеристикою — навіть технічно коректно працююча система втрачає практичну цінність, якщо її записи неточні, неповні або суперечливі. Зовнішні сервіси надають багаті, але неоднорідні метадані, тому система має підтримувати контроль повноти, точності, узгодженості та актуальності записів, а також фіксацію статусу верифікації окремо для кожного значення показника.

Вимога масштабованості (NFR-09) безпосередньо пов'язана з обраним архітектурним підходом до моделі даних. Якщо система зберігає метрики через довідник типів показників (MetricTypes) і узагальнену таблицю значень (MetricValues), як описано в підрозділі 3.3, додавання нової платформи зводиться до поповнення довідника Platforms, нової метрики — до поповнення MetricTypes, без зміни структури центральних таблиць. Це задає реалістичний цільовий показник: підтримка зростання до 10^6 записів значень показників і розширення набору сутностей без структурної перебудови бази даних. Вимога сумісності (NFR-05) забезпечує здатність системи працювати у спільному цифровому просторі з ORCID, Crossref, OpenAlex та іншими інфраструктурами через стандартні ідентифікатори та формати обміну даними [13; 26; 28].

Вимоги супроводжуваності (NFR-06) та зручності використання (NFR-07) визначають довготривалу перспективу розвитку системи. Зовнішні наукові сервіси регулярно змінюють свої API, платформи розширюють набір метаданих, а університет формує нові аналітичні запити, тому архітектура має передбачати модульність кодової бази та чіткий поділ між шарами роботи з даними, інтеграції та представлення. Зручність використання реалізується через диференціацію інтерфейсу за трьома базовими сценаріями — персональним (науковець переглядає та уточнює власний профіль), управлінським (адміністратор підрозділу або аналітик працює з вибірками й рейтингами) та системним (адміністратор керує довідниками й верифікацією).

Сформовані нефункціональні вимоги NFR-01...NFR-09 разом із функціональними вимогами FR-01...FR-11 (підрозділ 2.3) утворюють повну специфікацію вимог до інформаційної системи обліку наукометричної активності. Ця специфікація задає рамки для подальшої розробки моделі даних — визначення сутностей предметної області, побудови концептуальної та логічної моделей, що становить зміст третього розділу роботи.

Висновки до розділу 2

У другому розділі проаналізовано практичний контекст функціонування сервісу publication.ksru.edu, обґрунтовано напрями його розвитку та сформовано специфікацію функціональних і нефункціональних вимог до інформаційної системи обліку наукометричної активності. Це відповідає другому, третьому та четвертому завданням роботи, визначеним у вступі.

Аналіз публічно доступних розділів сервісу publication.kspu.edu — рейтингів факультетів, кафедр і науково-педагогічних працівників, аналітики, документів за галузями знань — показав, що наявне рішення вже реалізує багаторівневе представлення наукової активності та персональні сторінки науковців. Водночас встановлено, що архітектурні обмеження поточного стану пов'язані не з відсутністю інформаційної основи як такої, а з потребою її подальшої формалізації та модернізації. Попередня реалізація publication.kspu.edu мала базу даних, SQL-обробку, JSON-обмін, API-доступ і аналітичні модулі, однак подальший розвиток сервісу як сучасної CRIS-орієнтованої системи потребує формалізованої моделі даних, системної підтримки історичних зрізів, уніфікованого представлення різномірних показників, фіксації джерел походження та статусів верифікації даних.

Обґрунтовано напрями подальшого розвитку: розширення спектра джерел через API-інтеграцію з Crossref, OpenAlex та ORCID, перехід до узагальненого представлення показників на основі довідникової моделі типів метрик, збереження історичних зрізів даних, посилення ролі стійких цифрових ідентифікаторів (ORCID, DOI, ROR, ISSN) та поєднання зовнішніх наукометричних даних із внутрішнім університетським контекстом.

На основі сформованих напрямів розроблено специфікацію з 11 функціональних вимог (FR-01...FR-11, таблиця 2.2), згрупованих у чотири логічні блоки: облік базових сутностей предметної області, інтеграція із зовнішніми джерелами, аналітика й часовий контекст, керованість і доступ. Для деталізації функціональної вимоги керування доступом (FR-11) розроблено матрицю прав доступу за моделлю CRUD (Додаток А), що регламентує повноваження чотирьох базових ролей: науковця, адміністратора підрозділу або аналітика, системного адміністратора та зовнішнього API.

Сформовано специфікацію з 9 нефункціональних вимог (NFR-01...NFR-09, таблиця 2.3) на основі моделі якості програмних продуктів ISO/IEC 25010, моделі якості даних ISO/IEC 25012 та рекомендацій OWASP щодо захищеності прикладних систем. Визначено цільові показники для кожної вимоги, зокрема час відгуку інтерфейсу до 2 секунд для типових операцій, час виконання агрегованих аналітичних запитів до 5 секунд, підтримка зростання обсягу записів до 10^6 , розмежування доступу між чотирма ролями користувачів та обов'язкова фіксація статусу верифікації для кожного значення показника.

Сформована в другому розділі специфікація вимог (FR-01...FR-11 та NFR-01...NFR-09) задає рамки для подальшої розробки моделі даних: визначення основних сутностей предметної області, побудови концептуальної ER-моделі та логічної реляційної моделі бази даних, що становить зміст третього розділу роботи.

РОЗДІЛ 3 РОЗРОБКА МОДЕЛЕЙ ДАНИХ ІНФОРМАЦІЙНОЇ СИСТЕМИ ОБЛІКУ НАУКОМЕТРИЧНОЇ АКТИВНОСТІ

3.1 Визначення основних сутностей предметної області

Побудову моделі даних інформаційної системи обліку наукометричної активності варто починати не з переліку таблиць або технічних полів, а з чіткого окреслення сутностей предметної області. Саме на цьому етапі закладається логіка майбутньої системи: які об'єкти вона повинна відображати, які зв'язки між ними є змістовно необхідними, а які атрибути мають визначальне значення для обліку, аналітики та подальшого розвитку. Для теми цієї роботи це особливо важливо, адже наукометрична активність не зводиться до однорідного набору числових показників. Вона формується як результат взаємодії дослідника, його публікацій, профілів у зовнішніх платформах, афіліацій, структурних підрозділів, цифрових ідентифікаторів, джерел метаданих і власне наукометричних показників, що змінюються з часом. Отже, предметна область системи має багатосутнісний характер і не може бути коректно описана спрощеною схемою «автор — перелік публікацій».

Центральною сутністю предметної області є науковець. Саме навколо нього в університетській системі обліку зосереджується значна частина інформації: персональні дані, афіліації, належність до структурного підрозділу, авторські профілі, перелік наукових результатів, зовнішні ідентифікатори та пов'язані показники. На відміну від абстрактного автора публікації, науковець у цій системі виступає внутрішнім університетським об'єктом, що має організаційний контекст і бере участь у внутрішній аналітиці. Тому ця сутність має охоплювати не лише ідентифікаційні дані особи, а й зв'язки з підрозділами, посадами, періодами афіліації та зовнішніми профілями. У практичному вимірі це дає змогу поєднати внутрішній кадрово-організаційний

контекст університету із зовнішнім цифровим представленням наукової активності.

Поряд із центральною сутністю науковця потрібно враховувати і профіль науковця. Потреба в його виокремленні пов'язана з тим, що одна й та сама особа може бути представлена в кількох зовнішніх системах, кожна з яких має власну логіку формування профілю, власні ідентифікатори та набір доступних даних. Тому в моделі даних доцільно розмежувати сутність «науковець» як внутрішній об'єкт системи та сутність «профіль науковця» як форму представлення цієї особи в конкретному зовнішньому середовищі. Таке розмежування є принциповим, оскільки дає змогу працювати з кількома профілями однієї особи, не змішуючи їх із внутрішньою університетською карткою працівника.

Наступною базовою сутністю є публікація або, ширше, наукова робота. Саме навколо неї зосереджена значна частина бібліографічних, цитатних і аналітичних даних. Для інформаційної системи це означає, що публікація має бути повноцінним інформаційним об'єктом, який містить не лише назву чи бібліографічний опис, а й тип документа, цифрові ідентифікатори, джерело публікації, авторський склад та інші пов'язані характеристики. Саме ця сутність виступає основною ланкою між науковцем, платформою, джерелом даних і показниками наукометричної активності.

Водночас доцільно розмежувати сутність публікації та сутність джерела публікації. У контексті наукометричної активності важливо відрізнити саму роботу від того середовища, де вона була оприлюднена: журналу, збірника, конференції, репозитарію чи іншого ресурсу. Для майбутньої інформаційної системи це означає, що джерело публікації не повинно зводитися до звичайного текстового поля в бібліографічному записі, оскільки воно може мати власні характеристики, зокрема назву, тип, ISSN, видавця, платформену представленість або пов'язані

журнальні метрики. Виокремлення цієї сутності важливе і для аналітики, і для подальшого масштабування системи.

Окремого виділення потребує сутність наукометрична платформа. Необхідність її введення впливає з того, що одна й та сама публікація або один і той самий науковець можуть бути представлені в різних інформаційних системах, а самі показники набувають змісту лише разом із посиланням на платформу-джерело. Тому в моделі даних платформа має бути окремою сутністю, яка дозволяє описати середовище походження профілю, метаданих або показників. Це особливо важливо для системи, що повинна працювати з багатьма зовнішніми джерелами одночасно і уникати методичної помилки, коли значення показника зберігається без зазначення середовища, у якому воно було отримане. При цьому наукометричну платформу не слід ототожнювати з джерелом публікації: перша описує середовище індексування або представлення даних, друга — середовище оприлюднення самої роботи.

Ключовою для теми роботи є сутність наукометричний показник. Проте її слід розглядати не як просте поле типу «кількість цитувань» або «h-index», а як окремий інформаційний об'єкт. Такий об'єкт повинен мати власну назву, тип, рівень застосування, джерело та логіку інтерпретації. Це зумовлено тим, що різні показники можуть характеризувати різні об'єкти: автора, публікацію, журнал, інституцію чи іншу сутність. Тому система повинна розрізняти показник як тип аналітичної характеристики і значення показника як конкретний зафіксований факт.

Із цього безпосередньо впливає потреба у виокремленні сутності значення показника. Якщо показник є узагальненим типом метрики, то значення показника — це конкретний запис про його числове або інше значення щодо певного об'єкта, на певній платформі та на певну дату. Наприклад, кількість цитувань не є сталою характеристикою: вона змінюється з часом і може відрізнятися залежно від платформи. Саме

тому в моделі даних важливо відділяти те, що вимірюється, від того, яке саме значення було зафіксоване. Такий підхід дає змогу будувати часові зрізи, порівнювати дані між платформами та уникати жорсткого кодування кожної метрики як окремого стовпця бази даних.

Суттєве значення для системи має також сутність афіліація. У наукометричному середовищі зв'язок дослідника з установою є не менш важливим, ніж сам факт наявності публікації. Для університетської системи ця сутність дозволяє не лише зберігати факт належності науковця до закладу, а й враховувати зміни афіліації в часі, що особливо важливо для коректної аналітики на рівні підрозділів та інституції загалом. У такому разі афіліація виступає не просто атрибутом особи, а окремим змістовним зв'язком між науковцем та організацією.

З цією логікою пов'язана сутність структурний підрозділ. Для цілей університетського обліку наукометричної активності система повинна працювати не лише з окремими науковцями, а й із кафедрами, факультетами, інститутами або іншими організаційними одиницями. Це впливає і з практики publication.kspu.edu, де аналітика представлена на рівні факультетів, кафедр та інших структурних розрізів. Отже, структурний підрозділ має бути окремою сутністю, а не просто довідковим текстом у картці працівника, оскільки саме на його основі формуються рейтинги, агреговані звіти та аналітичні зрізи. Для моделі даних це означає, що підрозділи повинні бути включені в систему зв'язків нарівні з науковцями, публікаціями та показниками.

Окремо варто визначити сутність зовнішній ідентифікатор. Хоча технічно ORCID iD, DOI чи ROR можуть зберігатися як атрибути відповідних об'єктів, з погляду гнучкої моделі даних доцільніше розглядати їх як окремий тип пов'язаної інформації. Це зумовлено тим, що один об'єкт може мати кілька ідентифікаторів різних типів, а самі ці ідентифікатори виконують роль ключів зв'язування системи із зовнішніми сервісами. Тому в моделі даних доцільно передбачити

механізм системного зберігання зовнішніх ідентифікаторів із зазначенням їх типу, значення, джерела та об'єкта, до якого вони належать.

Ще однією важливою сутністю є джерело даних або походження даних. У системі обліку наукометричної активності недостатньо знати саме значення показника чи метаданих; потрібно розуміти, звідки це значення походить, яким способом воно було отримане, коли саме було імпортоване або підтверджене. Це означає, що для прозорості, верифікованості та подальшої аналітики система повинна фіксувати джерело даних окремо від самих сутностей предметної області. Такий підхід дозволяє розрізняти записи, введені вручну, імпортовані з API, уточнені користувачем або синхронізовані з конкретною платформою.

Для коректного обліку динамічних характеристик важливою є також сутність часовий зріз або момент фіксації. Значна частина наукометричних даних змінюється: зростає кількість цитувань, уточнюються профілі, змінюються афіліації, доповнюються зовнішні записи. Тому система повинна мати можливість фіксувати не лише сам факт наявності показника або запису, а й його стан у конкретний момент часу. У моделі даних ця сутність може бути представлена як дата фіксації, дата оновлення, період дії або інший часовий атрибут, однак концептуально її слід розуміти як окремий вимір обліку. Без цього система не зможе підтримувати повноцінну історичну аналітику та коректно працювати з динамікою наукометричної активності.

Крім того, для реального функціонування системи слід врахувати, що публікації можуть мати не лише внутрішніх університетських авторів, а й зовнішніх співавторів. Тому концепція предметної області повинна бути достатньо гнучкою, щоб поєднувати внутрішній об'єкт «науковець» із ширшим колом авторів, які беруть участь у створенні наукових робіт, але не обов'язково належать до кадрового складу університету. Такий підхід підвищує адекватність моделі реальній

науковій комунікації та запобігає штучному звуженню предметної області лише до внутрішніх учасників.

Отже, у межах предметної області інформаційної системи обліку наукометричної активності до ключових сутностей доцільно віднести науковця, профіль науковця, публікацію, джерело публікації, наукометричну платформу, наукометричний показник, значення показника, афіліацію, структурний підрозділ, зовнішній ідентифікатор, джерело даних і часовий зріз. Саме така система сутностей формує концептуальну основу для подальшої побудови моделі зв'язків, а згодом — і логічної моделі даних, орієнтованої на вимоги системи та особливості наукометричної активності як об'єкта обліку.

3.2 Розробка концептуальної моделі даних

Після визначення основних сутностей предметної області наступним кроком є побудова концептуальної моделі даних. Саме вона дає змогу перейти від опису окремих об'єктів до цілісного бачення інформаційної системи, у межах якого стає зрозуміло, які сутності є центральними, як вони пов'язані між собою, які зв'язки мають базове значення, а які виконують допоміжну або уточнювальну роль. Для теми цієї роботи це має особливе значення, оскільки інформаційна система обліку наукометричної активності повинна відображати не лише сам факт наявності публікацій або показників, а всю систему взаємозв'язків між науковцем, його профілями, публікаційною активністю, платформами, джерелами даних, структурними підрозділами та аналітичними характеристиками.

У межах цієї роботи концептуальна модель розглядається як узагальнене подання предметної області, незалежне від конкретної системи керування базами даних або технічної реалізації. Її головне призначення полягає у формалізації змістових зв'язків між об'єктами, які були виявлені в попередніх розділах, і створенні підґрунтя для

подальшого переходу до логічної моделі. Якщо в першому розділі було показано, що наукометрична активність є складною, динамічною та багатоджерельною предметною областю, а в другому визначено вимоги до системи, то саме концептуальна модель має продемонструвати, як ці вимоги можуть бути відображені у структурі даних.

Центральне місце в запропонованій концептуальній моделі займає сутність «Науковець». Це зумовлено тим, що саме науковець виступає основним суб'єктом наукометричної активності в межах університетської системи. З ним пов'язані публікації, авторські профілі, зовнішні ідентифікатори, афіліації, структурні підрозділи та частина аналітичних показників. У контексті цієї роботи науковець розглядається не лише як автор публікації, а як внутрішній об'єкт інформаційної системи університету, тобто як особа, щодо якої формуються облік, аналітика та управлінські представлення.

Із сутністю «Науковець» безпосередньо пов'язана сутність «Профіль науковця». Логіка такого зв'язку полягає в тому, що один науковець може мати кілька профілів у різних зовнішніх системах. Відповідно, між сутностями «Науковець» і «Профіль науковця» доцільно встановити зв'язок один-до-багатьох. Один науковець може бути представлений у кількох зовнішніх середовищах, тоді як кожен конкретний профіль належить лише одному науковцю в межах внутрішньої моделі системи. Такий підхід дає змогу коректно враховувати різні зовнішні представлення однієї особи без дублювання внутрішньої сутності працівника.

Сутність «Профіль науковця», своєю чергою, пов'язується із сутністю «Наукометрична платформа». У цьому випадку доцільним є зв'язок багато-до-одного з боку профілю або, відповідно, один-до-багатьох з боку платформи. Одна платформа може містити велику кількість профілів різних науковців, тоді як кожен конкретний профіль формується в межах лише однієї платформи. Такий зв'язок важливий не

тільки для ідентифікації джерела профілю, а й для подальшої інтерпретації показників, оскільки значення метрик мають сенс лише разом із зазначенням середовища, у якому вони були отримані.

Наступним ключовим елементом концептуальної моделі є сутність «Публікація». Вона пов'язана з науковцем через зв'язок багато-до-багатьох, оскільки один науковець може бути автором багатьох публікацій, а одна публікація може мати кількох авторів. Саме цей зв'язок є одним із найважливіших у всій моделі, бо відображає реальну природу наукової комунікації, де співавторство є типовим явищем. На концептуальному рівні це означає, що в подальшій логічній моделі має бути передбачено окрему проміжну сутність або асоціативну таблицю, яка забезпечить коректне представлення співавторства, порядку авторів, ролей або інших характеристик авторської участі.

Поряд із цим важливо врахувати, що не всі автори публікацій обов'язково є внутрішніми об'єктами університетської системи. У реальній науковій комунікації значна частина робіт створюється у співавторстві із зовнішніми дослідниками. Тому концептуальна модель повинна бути достатньо гнучкою, щоб поєднувати внутрішній об'єкт «Науковець» із ширшим колом авторів, які беруть участь у створенні наукових робіт, але не належать до кадрового складу університету. Це дає змогу уникнути штучного звуження предметної області та робить модель ближчою до реальної практики наукової діяльності.

Із сутністю «Публікація» пов'язана сутність «Джерело публікації». У цьому випадку доцільно встановити зв'язок багато-до-одного з боку публікації або один-до-багатьох з боку джерела публікації. Одне джерело, наприклад журнал або збірник матеріалів конференції, може містити багато публікацій, тоді як конкретна публікація у звичайному випадку пов'язується з одним основним джерелом оприлюднення. Такий зв'язок має значення для подальшого аналізу публікаційної активності

за журналами, збірниками, конференційними майданчиками чи іншими видами джерел.

Сутність «Наукометричний показник» у запропонованій концептуальній моделі не прив'язується лише до одного типу об'єктів. Її особливість полягає в тому, що один і той самий тип показника може застосовуватися до різних сутностей: науковця, публікації, джерела публікації або структурного підрозділу. Саме тому на концептуальному рівні доцільно розмежувати два рівні: показник як узагальнений тип аналітичної характеристики і «Значення показника» як конкретний факт його фіксації. У такому разі між сутностями «Наукометричний показник» і «Значення показника» встановлюється зв'язок один-до-багатьох, оскільки один тип показника може мати багато конкретних значень, зафіксованих для різних об'єктів, на різних платформах і в різні моменти часу.

Сутність «Значення показника» є однією з найважливіших для побудови гнучкої моделі. Вона виконує роль зв'язувальної ланки між метрикою, об'єктом вимірювання, платформою, джерелом даних і часовим зрізом. Саме завдяки цій сутності система може зберігати не просто число, а контекстно пов'язаний запис: що саме вимірювалося, для якого об'єкта, де було отримано значення і коли воно було зафіксоване. У концептуальній моделі це означає, що «Значення показника» повинно мати зв'язки принаймні з такими сутностями, як «Наукометричний показник», «Наукометрична платформа», «Джерело даних» і «Часовий зріз». За логікою моделі воно також пов'язується з конкретним об'єктом вимірювання — науковцем, публікацією, джерелом публікації або структурним підрозділом.

Окреме місце в моделі займають сутності «Афіліація» та «Структурний підрозділ». Їх виокремлення пов'язане з тим, що облік наукометричної активності в університетському середовищі не обмежується рівнем окремої особи. Система повинна підтримувати

аналітику за кафедрами, факультетами, інститутами чи іншими організаційними одиницями. У зв'язку з цим між сутностями «Науковець» і «Структурний підрозділ» доцільно не встановлювати спрощений прямий зв'язок, а опосередковувати його через сутність «Афіліація». Це пояснюється тим, що афіліація відображає не лише сам факт належності, а й контекст цього зв'язку, зокрема його часові межі, тип або рівень. Відповідно, один науковець може мати кілька афіліацій у різні періоди або навіть одночасно, а один структурний підрозділ може бути пов'язаний із багатьма науковцями. Отже, між «Науковцем» і «Структурним підрозділом» концептуально реалізується зв'язок багато-до-багатьох через сутність «Афіліація».

Сутність «Зовнішній ідентифікатор» у концептуальній моделі доцільно розглядати як універсальний механізм зв'язування внутрішніх об'єктів системи із зовнішніми цифровими середовищами. Один науковець може мати ORCID, один профіль — платформенний ідентифікатор, одна публікація — DOI, одне джерело публікації — ISSN або інший стандартний код, одна організація — ROR чи інший інституційний ідентифікатор. Саме тому сутність «Зовнішній ідентифікатор» доцільно пов'язувати не лише з одним класом об'єктів, а з різними сутностями предметної області. На концептуальному рівні це означає, що система повинна підтримувати механізм закріплення одного або кількох зовнішніх ідентифікаторів за різними типами внутрішніх об'єктів. Таке рішення підвищує гнучкість моделі та спрощує інтеграцію з API зовнішніх платформ.

Сутність «Джерело даних» відображає походження конкретних записів. Це особливо важливо для системи, що працює з поєднанням внутрішніх і зовнішніх даних. Один запис може бути введений вручну, інший — імпортований з API, третій — уточнений після верифікації користувачем. У концептуальній моделі сутність «Джерело даних» доцільно пов'язувати насамперед із сутністю «Значення показника», а за

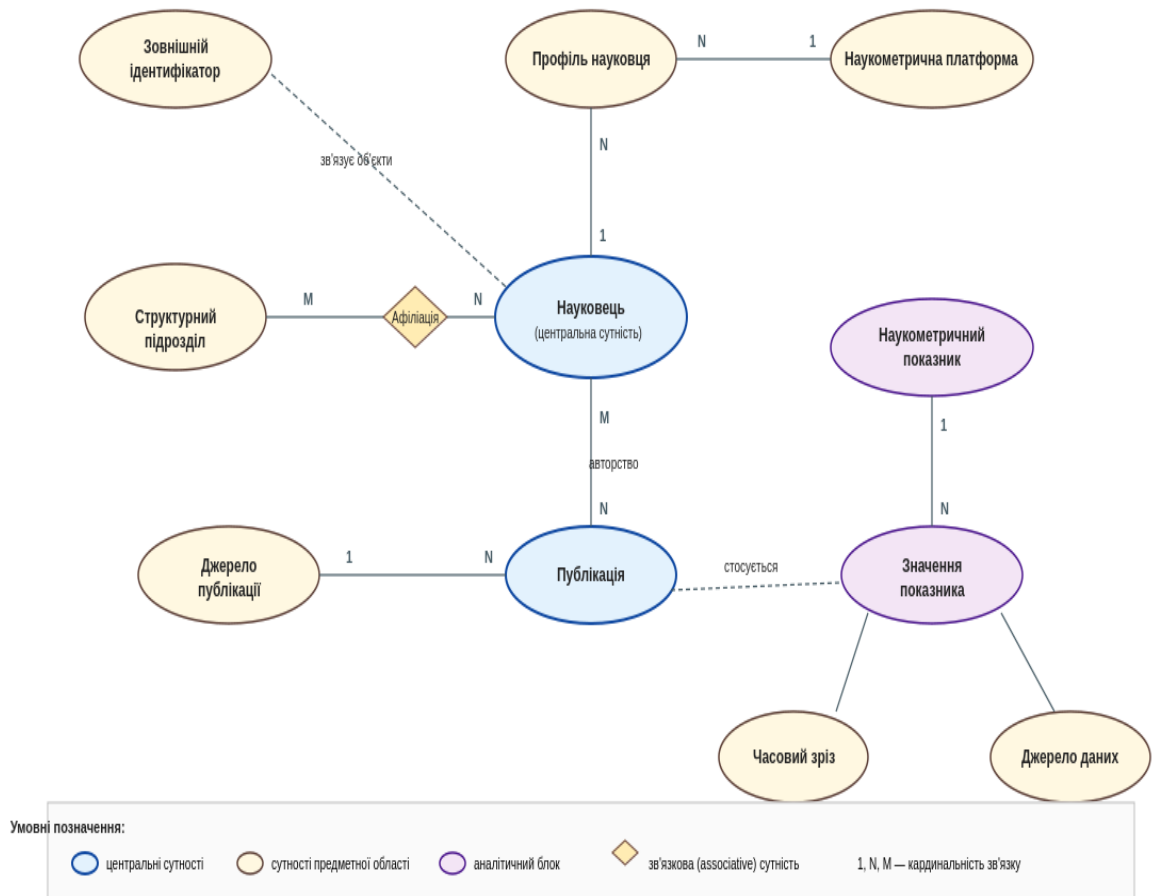
потреби — також із профілем, публікацією чи іншими об'єктами. Основна ідея полягає в тому, що походження інформації є окремим аспектом обліку, який не повинен губитися серед основних даних.

Не менш важливою є сутність «Часовий зріз». Оскільки наукометричні дані є динамічними, система повинна мати концептуальний механізм фіксації моменту або періоду, до якого належить певне значення. Один і той самий показник для того самого об'єкта може мати різні значення на різні дати. Тому зв'язок «Значення показника» — «Часовий зріз» є принципово важливим для підтримки історичної аналітики, порівняння даних і відстеження динаміки наукової активності. На концептуальному рівні часовий зріз виступає окремим виміром обліку, навіть якщо в подальшій логічній моделі він буде реалізований через спеціалізовані атрибути або окрему таблицю.

Узагальнюючи запропоновану концептуальну модель, можна виділити кілька її основних логічних центрів. Перший утворюють науковець, профілі, афіліації та структурні підрозділи, тобто блок ідентифікації суб'єкта та його організаційного контексту. Другий формують публікація та джерело публікації, тобто блок представлення власне наукового результату. Третій складають платформи, показники, значення показників, джерела даних і часові зрізи — блоки, пов'язані з аналітикою, вимірюванням та походженням інформації. Саме така структура дозволяє уникнути спрощення предметної області й водночас забезпечує достатню логічну впорядкованість для подальшого переходу до логічної моделі.

Концептуальну модель даних предметної області у вигляді ER-діаграми наведено на рисунку 3.1.

Рисунок 3.1 — Концептуальна модель даних (ER-діаграма)
інформаційної системи обліку наукометричної активності



3.3 Розробка логічної моделі даних

Після побудови концептуальної моделі наступним етапом є розробка логічної моделі даних, тобто такого формалізованого подання предметної області, у якому концептуальні сутності та зв'язки переводяться у структуру таблиць, атрибутів, ключів і обмежень цілісності. На відміну від концептуальної моделі, логічна модель вже орієнтується на правила реляційного подання даних, однак ще не прив'язується жорстко до конкретної системи керування базами даних чи особливостей фізичного розміщення інформації. Для цієї дипломної

роботи логічна модель має особливе значення, оскільки саме вона показує, як вимоги до обліку наукометричної активності можуть бути реалізовані в структурі даних, придатній до масштабування, інтеграції із зовнішніми джерелами та побудови аналітичних представлень.

Логічну модель доцільно будувати навколо кількох базових таблиць, які відображають головні сутності предметної області. Центральною таблицею виступає Scientists, або в українському описі — таблиця «Науковці». Її основне призначення полягає у збереженні внутрішніх університетських даних про особу, яка є суб'єктом наукометричної активності. На рівні логічної моделі ця таблиця повинна містити первинний ключ scientist_id, а також набір основних атрибутів, зокрема прізвище, ім'я, по батькові, варіанти написання імені, службовий статус, внутрішній ідентифікатор у межах університету та інші характеристики, релевантні для внутрішнього обліку. Важливо, що ця таблиця не повинна дублювати зовнішні профілі, оскільки одна й та сама особа може бути представлена в кількох зовнішніх системах. Саме тому внутрішній запис науковця виконує роль стабільної опорної сутності, з якою пов'язуються інші елементи моделі.

Для представлення зовнішніх середовищ доцільно ввести таблицю Platforms — «Платформи». Вона повинна містити довідкові записи про наукометричні системи або цифрові джерела представлення наукової активності, з якими працює інформаційна система. Первинним ключем цієї таблиці може бути platform_id, а серед основних атрибутів варто передбачити назву платформи, її тип, короткий опис, базову URL-адресу, ознаку підтримки API та інші характеристики, потрібні для інтеграції. Виокремлення цієї таблиці є логічно виправданим, оскільки одна платформа може бути джерелом багатьох профілів, багатьох значень показників і зовнішніх записів, але в самій системі повинна бути описана лише один раз.

Окрему таблицю повинна утворювати сутність `ScientistProfiles` — «Профілі науковців». У логічній моделі ця таблиця є зв'язувальною між внутрішнім науковцем і конкретною платформою. Її первинним ключем може бути `profile_id`, а зовнішніми ключами — `scientist_id` і `platform_id`. Серед атрибутів доцільно передбачити платформений ідентифікатор профілю, URL профілю, статус активності, дату першої реєстрації в системі, дату останнього оновлення, а також, за потреби, поле для приміток щодо якості або повноти профілю. Така структура дозволяє реалізувати зв'язок «один науковець — багато профілів» без дублювання персональних записів.

Для збереження бібліографічної інформації потрібна таблиця `Publications` — «Публікації». Саме вона повинна виступати центральним об'єктом для опису наукових результатів. Первинним ключем тут може бути `publication_id`. Серед основних атрибутів доцільно закласти назву публікації, рік видання, тип документа, мову, DOI, інші зовнішні ідентифікатори, короткий бібліографічний опис, статус верифікації, а також зовнішній ключ `source_id`, який посилається на джерело публікації. Для університетської системи це означає, що публікація повинна бути не просто текстовим записом, а повноцінним логічним об'єктом з набором атрибутів і зв'язків.

Оскільки одна публікація може бути пов'язана з багатьма авторами, а один автор — з багатьма публікаціями, прямого зовнішнього ключа між таблицями «Науковці» і «Публікації» недостатньо. Саме тому в логічній моделі необхідно ввести асоціативну таблицю `PublicationAuthors` — «Авторство публікацій». Її основне призначення — реалізація зв'язку багато-до-багатьох між науковцями та публікаціями. Доцільно або використати складений первинний ключ із полів `publication_id` і `scientist_id`, або ввести окремий сурогатний ключ `publication_author_id`. Крім зовнішніх ключів на таблиці «Публікації» та «Науковці», ця таблиця може містити атрибути `author_order`,

is_corresponding_author, role_type та інші характеристики, які дають змогу точніше відобразити авторську участь. Саме ця таблиця створює логічну основу для коректного представлення співавторства.

Для опису середовища оприлюднення доцільно запровадити таблицю Sources — «Джерела публікацій». Її первинним ключем може бути source_id. Вона повинна містити назву джерела, тип джерела, ISSN або ISSN-L, видавця, країну, вебпосилання та інші атрибути, що мають аналітичне значення. Саме на цю таблицю має посылатися таблиця «Публікації». Її виокремлення дає змогу уникнути дублювання інформації про журнал чи конференцію в кожному бібліографічному записі та відкриває можливість подальшої аналітики за джерелами.

Для роботи з аналітичними характеристиками логічна модель повинна містити таблицю MetricTypes — «Типи показників». Її роль полягає в тому, щоб зберігати не конкретні значення, а самі типи метрик. Первинним ключем тут може бути metric_type_id. До основних атрибутів доцільно віднести назву показника, опис, рівень застосування, тип значення, одиницю вимірювання, логічну групу та ознаку активності. Саме ця таблиця забезпечує уніфіковане представлення різноманітних наукометричних показників у межах єдиної структури. Замість створення окремого стовпця під кожен метрику система зберігає тип показника як самостійну сутність, а конкретні значення — окремо. Такий підхід є принципово важливим для масштабованості моделі. Фізичну структуру та обмеження ключової таблиці MetricValues («Значення показників»), що забезпечує гнучкість системи, наведено в таблиці 3.1 [17; 20; 23; 24; 50].

Таблиця 3.1

Структура атрибутів логічної таблиці MetricValues

Назва поля (атрибут)	Тип даних (SQL)	Обмеження	Призначення в системі
metric_value_id	INT / SERIAL	Primary Key (PK)	Унікальний ідентифікатор конкретного запису значення
metric_type_id	INT	Foreign Key (FK)	Посилання на тип показника (h-

Назва поля (атрибут)	Тип даних (SQL)	Обмеження	Призначення в системі
			index, цитування тощо)
platform_id	INT	Foreign Key (FK)	Посилання на платформу-джерело (Scopus, Scholar)
data_source_id	INT	Foreign Key (FK)	Посилання на спосіб отримання (API, ручне введення)
snapshot_id	INT	Foreign Key (FK)	Посилання на часовий зріз або конкретну дату фіксації
numeric_value	FLOAT / REAL	Not Null	Власне числове значення наукометричного показника
object_type	VARCHAR	Not Null	Тип об'єкта застосування (Науковець, Публікація, Установа)
object_id	INT	Поліморфний ключ	Ідентифікатор об'єкта, якого стосується показник
verification_status	VARCHAR	CHECK / довідкове значення	Статус перевірки запису (Imported, Confirmed, Manual, Rejected)

На основі таблиці типів показників повинна функціонувати таблиця MetricValues — «Значення показників». Це одна з центральних таблиць логічної моделі, призначена для фіксації конкретних значень наукометричних показників щодо визначених об'єктів предметної області. Її первинним ключем є `metric_value_id`, а до складу основних зовнішніх ключів належать `metric_type_id`, `platform_id`, `data_source_id` та `snapshot_id`. Така структура дає змогу зберігати не лише саме значення показника, а й контекст його отримання: тип метрики, платформу-джерело, спосіб надходження даних та часовий зріз, до якого належить запис. Це відповідає загальній логіці представлення наукометричних метрик як структурованих об'єктів, пов'язаних із джерелом, об'єктом застосування та моментом фіксації, а не як ізольованих числових полів [17; 20; 23; 24; 26]. Додатково таблиця повинна містити поле `verification_status`, яке фіксує статус перевірки запису та дозволяє розмежовувати автоматично імпортовані, підтверджені, вручну скориговані або відхилені значення, що узгоджується з вимогами до якості та верифікованості даних у системі [35; 50].

Запропоноване рішення дає змогу зберігати в межах єдиної таблиці `MetricValues` значення показників для науковця, публікації, джерела публікації, структурного підрозділу та інших об'єктів без зміни базової структури таблиці. Основою такої гнучкості є поліморфний механізм зв'язування через атрибути `object_type` та `object_id`, який дозволяє прив'язувати запис до різних типів сутностей предметної області. Перевагою цього підходу є масштабованість моделі та можливість розширювати перелік об'єктів аналізу без структурної перебудови центральної таблиці. Водночас така побудова потребує додаткового контролю логічної цілісності на рівні прикладної логіки системи, оскільки зв'язок із конкретною сутністю визначається не окремим зовнішнім ключем, а комбінацією типу об'єкта та його ідентифікатора. Атрибут `snapshot_id` забезпечує зв'язок запису з відповідним часовим зрізом, що дає змогу відстежувати динаміку показників у часі, а поле `verification_status` фіксує ступінь перевіреності значення та відокремлює автоматично імпортовані записи від підтверджених або скоригованих. Саме така структура робить таблицю `MetricValues` придатною для накопичення історичних даних, побудови аналітичних порівнянь і подальшого масштабування системи [17; 20; 23; 24; 26; 50].

Для відображення організаційної структури університету потрібна таблиця `Units` — «Структурні підрозділи». Її первинним ключем може бути `unit_id`. Серед атрибутів доцільно передбачити назву підрозділу, тип підрозділу, код, батьківський підрозділ та інші характеристики, які дозволяють реалізувати ієрархію організаційної структури. Необхідність окремого зберігання таких сутностей впливає як із потреби внутрішньої університетської аналітики, так і з логіки зовнішніх наукових систем, де `institutions` подаються як окремі сутності.

Щоб коректно відобразити зв'язок між науковцем і підрозділом, доцільно ввести таблицю `Affiliations` — «Афіліації». Її первинним

ключем може бути `affiliation_id`. Вона повинна містити зовнішні ключі `scientist_id` і `unit_id`, а також атрибути, що описують контекст належності: тип афіліації, дату початку, дату завершення, основну чи додаткову роль, а також, за потреби, зовнішній ідентифікатор установи. Відокремлення афіліації в самостійну таблицю дає змогу відобразити часову мінливість і багатозначність організаційних зв'язків, не руйнуючи цілісності таблиці «Науковці».

Для забезпечення інтеграції із зовнішніми системами доцільно виокремити таблицю `ExternalIdentifiers` — «Зовнішні ідентифікатори». Вона повинна містити первинний ключ `external_identifier_id`, тип ідентифікатора, його значення, посилання на платформу, а також поле для посилання на об'єкт, до якого цей ідентифікатор належить. Як і у випадку зі значеннями показників, тут можливе або поліморфне посилання, або кілька спеціалізованих зовнішніх ключів. Змістовно ця таблиця потрібна для зберігання ORCID, DOI, ROR, ISSN-L та інших зовнішніх ключів, які виступають механізмом зв'язування з API і відкритими метаданими.

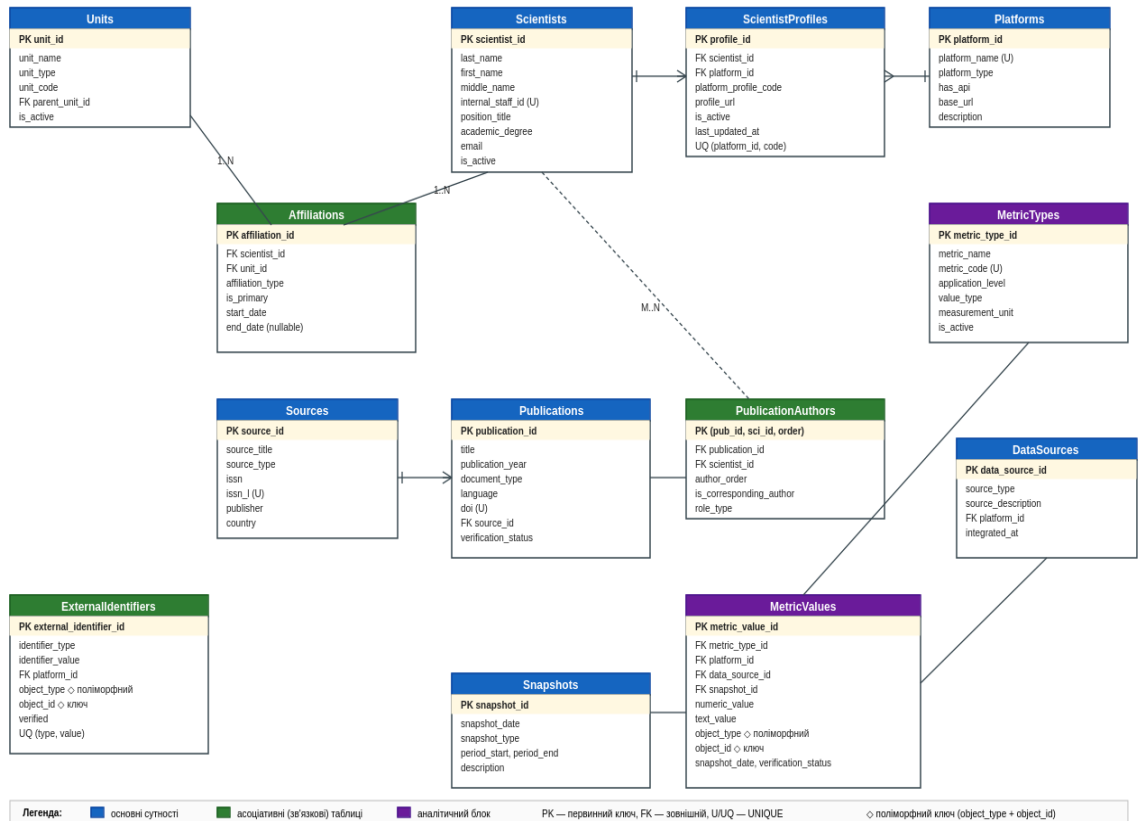
Для фіксації походження записів логічна модель повинна містити таблицю `DataSources` — «Джерела даних». Її первинним ключем може бути `data_source_id`. Вона повинна описувати, яким шляхом отримано запис: вручну, через API, шляхом імпорту з конкретної платформи, через верифікацію користувачем тощо. Серед атрибутів доцільно передбачити назву джерела, тип джерела, технічний опис, дату інтеграції або останнього звернення, а також, за потреби, параметри довіри чи статусу. Найприродніше пов'язувати цю таблицю зі значеннями показників, профілями та, за потреби, публікаціями. Така таблиця є особливо важливою, оскільки дозволяє фіксувати походження запису окремо від самого змісту цього запису.

Для підтримки історичності даних логічно передбачити таблицю `Snapshots` або `TimeSlices` — «Часові зрізи». У найпростішому варіанті її

роль можуть виконувати дати в таблиці значень показників, однак виділення окремої таблиці дозволяє системніше працювати з періодами, наборами імпортів або контрольними датами аналітики. Первинним ключем тут може бути `snapshot_id`, а серед атрибутів — дата фіксації, тип зрізу, опис періоду та інші часові характеристики. На цю таблицю доцільно посилати записи в таблиці «Значення показників». Така побудова дає можливість відстежувати динаміку змін і формувати порівняльну аналітику в часовому розрізі, що особливо важливо для цитувань, профільних даних і агрегованих рейтингів.

З точки зору логічної цілісності модель повинна спиратися на систему первинних і зовнішніх ключів, а також на обмеження унікальності та референтної цілісності. Наприклад, для таблиці профілів доцільно встановити обмеження, яке не дозволяє дублювати однаковий платформений ідентифікатор у межах однієї платформи; для зовнішніх ідентифікаторів — окремі правила унікальності залежно від типу ідентифікатора; для таблиці авторства — заборону повторного дублювання одного й того самого зв'язку між публікацією та науковцем. Такі обмеження особливо важливі в системах, що працюють з інтеграцією відкритих метаданих і стійкими цифровими ідентифікаторами. Детальну логічну ER-модель реляційної бази даних розробленої системи у нотації Crow's Foot зображено на рисунку 3.2.

Рисунок 3.2 — Логічна модель даних (Crow's Foot notation)



Для забезпечення інтегрованості системи із зовнішніми джерелами в межах роботи розроблено схему мапінгу вхідних JSON-метаданих від Crossref REST API у структуру запропонованої реляційної моделі (див. Додаток Б). У цій схемі враховано використання зовнішніх ідентифікаторів публікацій та авторів, зіставлення джерел публікацій, фіксацію способу надходження даних і прив'язку записів до відповідного часового зрізу. Практичну реалізацію центральних елементів розробленої логічної моделі подано у вигляді фрагмента DDL-скрипту створення таблиць `MetricTypes`, `Platforms`, `DataSources`, `Snapshots` і `MetricValues` мовою SQL (див. Додаток В). Це дозволяє показати перехід від концептуального та логічного рівнів моделювання до структур, придатних для подальшої програмної реалізації системи.

3.4 Узагальнення критеріїв представлення наукометричних показників у структурі даних системи

Розроблена в попередніх підрозділах модель даних має забезпечити роботу з різнорідними наукометричними метриками в узгодженому форматі. У реальному науковому середовищі показники не виникають у межах одного джерела і не мають єдиної, наперед заданої структури: одні платформи фокусуються на ідентифікації дослідника, інші — на бібліографічних метаданих, ще інші — на зв'язаних сутностях глобальної дослідницької системи. За таких умов інформаційна система не може будуватися за принципом жорсткого набору ізольованих полів під кожен окрему метрику. Натомість вона повинна спиратися на узагальнену модель представлення показників, здатну адаптуватися до змін зовнішнього середовища без структурної перебудови бази даних.

Потреба в узагальненні критеріїв пояснюється змістовою та структурною неоднорідністю самих метрик. Частина з них характеризує науковця (кількість публікацій, h-індекс), частина — окрему публікацію (кількість цитувань), інша частина належить до джерела публікації (CiteScore, SJR), установи або структурного підрозділу. Навіть подібні за назвою показники можуть мати різне значення залежно від платформи через відмінності в охопленні джерел, алгоритмах індексування та порядку оновлення даних. Звідси випливає, що зберігати лише числове значення і коротку назву показника недостатньо — потрібен набір універсальних характеристик, які забезпечують коректне тлумачення, порівняння та подальше використання метрики в аналітиці.

Сформовані в роботі критерії представлення наукометричного показника доцільно об'єднати у три логічні групи: критерії ідентифікації показника, критерії контексту вимірювання та критерії метайнформації запису. Така структура спрощує реалізацію в моделі даних і відображає різні аспекти опису метрики як інформаційного об'єкта.

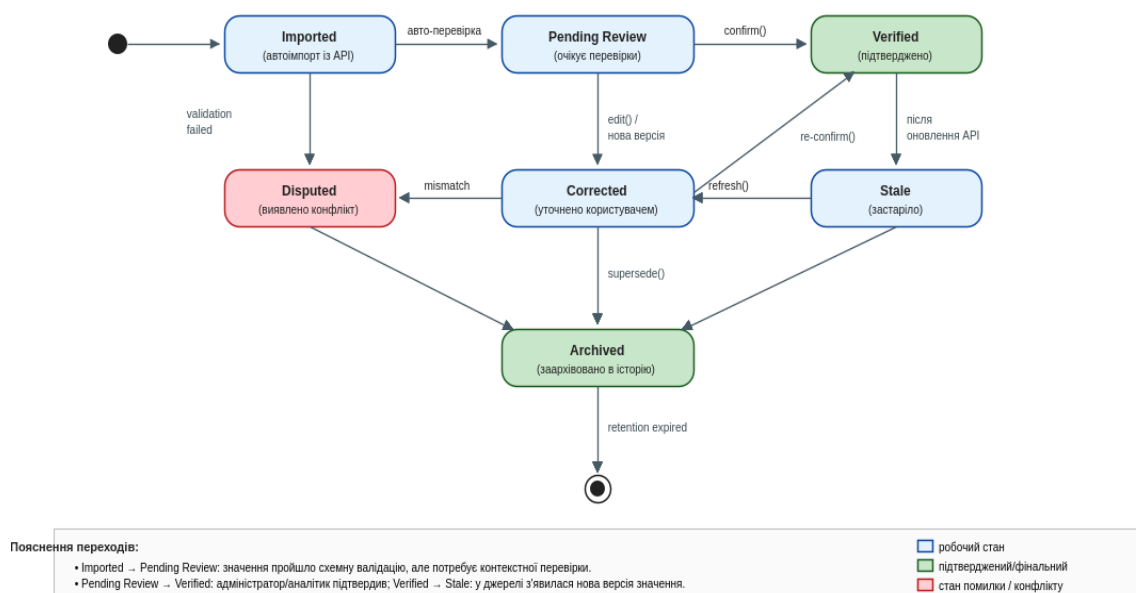
Критерії ідентифікації відповідають на питання «що саме вимірюється». До них належать назва показника — людиноорієнтована форма позначення метрики (наприклад, h-індекс, CiteScore, кількість цитувань); тип показника — класифікація за роллю в системі (кількісний, ранговий, бінарний, агрегований, похідний, ідентифікаційний); одиниця вимірювання або формат інтерпретації — характеристика, що формалізує спосіб подання значення (число, відсоток, ранг, категорія). Сама лише назва не є достатньою, оскільки в різних середовищах схожі назви можуть приховувати різний зміст: «cited by» у Google Scholar, «citations» у Scopus та «cited_by_count» в OpenAlex формально позначають одне явище, але обчислюються за різними правилами [23; 28; 32].

Критерії контексту вимірювання відповідають на питання «де, коли та для якого об'єкта зафіксовано показник». До них належать об'єкт застосування (науковець, профіль, публікація, джерело публікації, структурний підрозділ — кількість публікацій для автора і кількість публікацій для кафедри є різними фактами аналізу, хоча формально позначаються однаково), платформа-джерело (Scopus, Web of Science, Google Scholar, OpenAlex, ORCID — з якого середовища отримане значення) та часовий зріз (момент або період фіксації). Часовий критерій є особливо важливим для динамічних характеристик: цитування зростають із часом, профілі доповнюються новими роботами, агреговані індекси перераховуються. Без зв'язку показника з часом аналітика обмежується випадковим «знімком» і не дає змоги відстежувати динаміку [26; 32; 34].

Критерії метаінформації запису відповідають на питання «звідки походить значення і наскільки йому можна довіряти». Сюди належать джерело даних у вузькому сенсі (механізм отримання — ручне введення, API-імпорт, синхронізація, уточнення після перевірки) та статус верифікації (автоматично імпортоване, попереднє, підтверджене,

скориговане, потребує перевірки). Походження запису впливає на довіру до значення, спосіб його актуалізації та подальше використання в звітності й рейтингуванні; це критично для університетського середовища, де результати обліку можуть слугувати підставою для управлінських рішень [35]. Життєвий цикл статусу верифікації наукометричних даних та умови його зміни проілюстровано на рисунку 3.3.

Рисунок 3.3 — UML-діаграма станів: життєвий цикл верифікації наукометричних даних



Зведений перелік сформованих критеріїв з прикладами реалізації подано в таблиці 3.2.

Таблиця 3.2

Узагальнені критерії представлення наукометричного показника у структурі даних

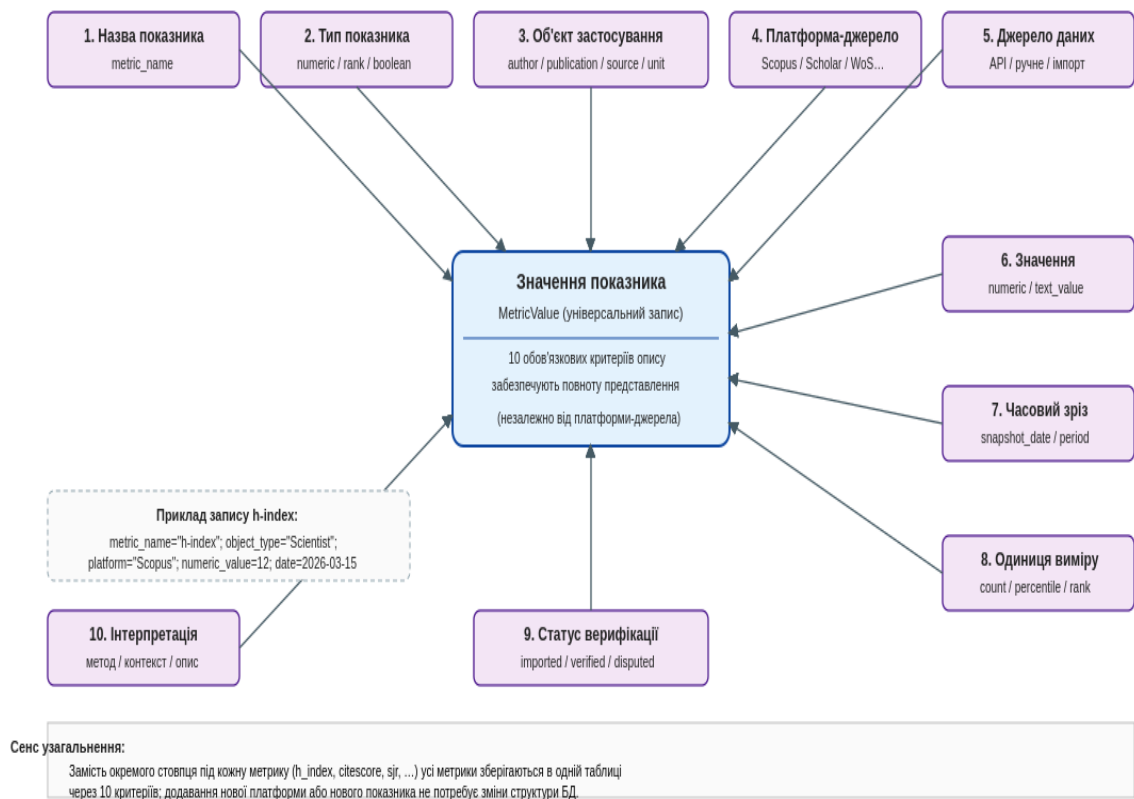
Група	Критерій	Призначення	Приклад значення	Реалізація в моделі
Ідентифікація	Назва показника	Людиноорієнтоване позначення метрики	h-індекс, CiteScore	MetricTypes.metric_name
Ідентифікація	Тип показника	Класифікація метрики за роллю	кількісний, ранговий	MetricTypes.metric_kind
Ідентифікація	Формат /	Спосіб	число,	MetricTypes.value

Група	Критерій	Призначення	Приклад значення	Реалізація в моделі
ія	одиниця	інтерпретації значення	ранг, відсоток	_format
Контекст	Об'єкт застосування	Сутність, до якої належить значення	Науковець, Публікація, Підрозділ	MetricValues.object_type + object_id
Контекст	Платформа-джерело	Середовище отримання даних	Scopus, OpenAlex	MetricValues.platform_id
Контекст	Часовий зріз	Момент або період фіксації	15.03.2026	MetricValues.snapshot_id / Snapshots.snapshot_date
Метаінформація	Джерело даних	Спосіб отримання запису	API-імпорт, ручне введення	MetricValues.data_source_id
Метаінформація	Статус верифікації	Ступінь довіри до запису	Imported, Confirmed, Manual	MetricValues.verification_status

Запропонований підхід доцільно проілюструвати на конкретному прикладі. Розглянемо фіксацію h-індексу науковця X, отриманого через Scopus API 15.03.2026 зі значенням 12. У структурі даних цей запис відображається не як ізольоване числове поле `h_index = 12` у профілі автора, а як комплексний об'єкт у таблиці `MetricValues`, що поєднує: тип метрики (h-індекс, кількісний, цілочисельний) → платформу-джерело (Scopus) → об'єкт застосування (Науковець, `scientist_id = N`) → часовий зріз (`snapshot_id`, що посилається на запис із датою фіксації 2026-03-15) → джерело даних (API-імпорт через Scopus REST) → статус (Imported, до уточнення). За такої побудови те ж саме значення h-індексу того самого автора з Google Scholar, отримане в інший час, фіксується в системі як окремий запис без переписування даних або зміни структури таблиць. Узагальнену структуру представлення наукометричного

показника як інформаційного об'єкта подано на рисунку 3.4.

Рисунок 3.4 — Узагальнена структура представлення наукометричного показника як інформаційного об'єкта



Запропоноване узагальнення має не лише технічний, а й методичний зміст: показник перестає бути ізольованим числом і перетворюється на контекстно пов'язаний аналітичний запис. Це робить модель придатною для подальшого розвитку *publication.kspu.edu*, оскільки додавання нових типів метрик, нових платформ або нових рівнів аналітики не потребує структурної зміни бази даних, а зводиться до поповнення довідників *MetricTypes*, *Platforms* та *DataSources*. Сформовані критерії створюють єдину логіку зберігання й інтерпретації різномірних метрик, забезпечують масштабованість системи, підвищують прозорість походження даних і слугують основою для побудови гнучкої аналітичної інфраструктури в університетському середовищі.

Висновки до розділу 3

У третьому розділі розроблено модель даних інформаційної системи обліку наукометричної активності на трьох взаємопов'язаних рівнях: на рівні визначення сутностей предметної області, на рівні концептуального моделювання зв'язків і на рівні логічного подання даних у структурі реляційної бази. Такий підхід дозволив перейти від загального опису вимог до формалізованої моделі, придатної для подальшої реалізації та розвитку системи.

У результаті дослідження визначено ключові сутності предметної області: науковець, профіль науковця, публікація, джерело публікації, наукометрична платформа, наукометричний показник, значення показника, афіліація, структурний підрозділ, зовнішній ідентифікатор, джерело даних і часовий зріз. Їх виокремлення показало, що наукометрична активність не може бути коректно подана у вигляді спрощеного набору полів або лінійного переліку публікацій, а потребує багатосутнісної моделі, яка враховує як внутрішній університетський контекст, так і взаємодію із зовнішніми цифровими платформами.

На основі визначених сутностей побудовано концептуальну модель даних, яка відображає змістові зв'язки між основними об'єктами системи. Установлено, що центральне місце в моделі посідає науковець як основний суб'єкт обліку, з яким пов'язані профілі, публікації, афіліації, структурні підрозділи та частина аналітичних характеристик. Концептуальна модель підтвердила, що предметна область має мережевий, динамічний і багаторівневий характер, а тому потребує формалізованої системи зв'язків, зокрема зв'язків типу «один до багатьох» і «багато до багатьох».

Подальшим етапом стала розробка логічної моделі даних, у межах якої концептуальні сутності було переведено у структуру таблиць, атрибутів, первинних і зовнішніх ключів. Запропонована логічна модель охоплює таблиці для науковців, платформ, профілів, публікацій, авторства, джерел публікацій, типів показників, значень показників, структурних підрозділів, афіліацій, зовнішніх ідентифікаторів, джерел даних і часових зрізів. Така структура забезпечує цілісність даних, підтримує інтеграцію із зовнішніми сервісами, дозволяє працювати з історичними значеннями та створює основу для подальшої аналітики.

Найважливішим результатом третього розділу стало обґрунтування узагальненого підходу до представлення наукометричних показників у структурі даних системи. Доведено, що для розвитку системи недоцільно зберігати кожен нову метрику як окреме ізольоване поле. Натомість запропоновано описувати показники через універсальний набір критеріїв, до яких належать назва, тип, об'єкт застосування, платформа-джерело, джерело даних, значення, часовий зріз, формат інтерпретації та статус верифікації. Саме такий підхід забезпечує масштабованість моделі, дозволяє додавати нові платформи та типи метрик без структурного руйнування бази даних і підвищує аналітичну придатність системи.

Отже, у третьому розділі сформовано повноцінну модельну основу інформаційної системи обліку наукометричної активності. Результати розділу підтверджують, що вимоги до розвитку такої системи можуть бути реалізовані через багатосутнісну, логічно впорядковану та масштабовану модель даних, орієнтовану на інтеграцію з сучасними цифровими науковими платформами, накопичення історичних даних і підтримку багаторівневої аналітики.

У бакалаврській роботі розв'язано актуальне прикладне завдання — формування вимог і розробку моделей даних для вдосконалення інформаційної системи обліку наукометричної активності на основі аналізу функціонування сервісу publication.kspu.edu. Отримані результати відповідають профілю спеціальності 126 «Інформаційні системи та технології», оскільки в роботі основну увагу зосереджено не на описі наукометрії як самодостатнього явища, а на системному аналізі предметної області, формалізації вимог та моделюванні даних для подальшого розвитку університетської інформаційної системи. Поставлену в роботі мету досягнуто, визначені завдання виконано в повному обсязі, що підтверджується наступними результатами.

За першим завданням досліджено наукометрію як предметну область інформаційної системи, схарактеризовано основні наукометричні платформи та визначено особливості наукометричної активності як об'єкта цифрового обліку. Розглянуто Scopus і Web of Science як комерційні реферативно-цитатні бази, Google Scholar як академічну пошукову платформу, ORCID як систему стійкої ідентифікації дослідників, Crossref і OpenAlex як відкриті інфраструктури метаданих, а також Dimensions як зв'язану аналітичну базу. Установлено, що ці системи надають дані різної природи й не можуть розглядатися як взаємозамінні джерела. Виявлено ключові проблеми автоматизації обліку — гетерогенність джерел, неоднорідність показників, складність ідентифікації автора та динамічний характер метрик; наголошено на важливості принципів відповідального

використання метрик, закріплених у Лейденському маніфесті та декларації DORA. Проведений порівняльний аналіз провідних рішень класу CRIS (Pure від Elsevier, відкритих DSpace-CRIS і VIVO) дозволив визначити архітектурні орієнтири для розвитку publication.kspu.edu — сумісність з моделлю CERIF, рекомендаціями OpenAIRE Guidelines for CRIS Managers, повну інтеграцію з ORCID API v3, підтримку часових зрізів і відкритих REST-інтерфейсів.

За другим завданням проаналізовано практичний контекст функціонування publication.kspu.edu, виявлено архітектурні обмеження поточної реалізації та обґрунтовано напрями розвитку системи. Сервіс вже виконує важливі функції в університетському цифровому середовищі — забезпечує персональні сторінки науковців, рейтинги факультетів, кафедр і науково-педагогічних працівників, відомості за галузями знань. Водночас встановлено, що подальший розвиток publication.kspu.edu потребує не створення інформаційної системи з нуля, а модернізації та формалізації її модельної основи: уточнення структури даних, уніфікації представлення показників, системної підтримки часових зрізів, розвитку API-орієнтованої інтеграції із зовнішніми джерелами та впровадження механізмів верифікації даних. До основних напрямів розвитку віднесено розширення спектра джерел даних, узагальнене представлення показників, посилення ролі цифрової ідентифікації дослідників, поєднання зовнішніх метрик із внутрішнім університетським контекстом і підвищення керованості й верифікованості інформації.

За третім завданням сформовано специфікацію функціональних вимог до системи, що містить 11 вимог (FR-01...FR-11, таблиця 2.1), згрупованих у чотири логічні блоки: облік базових сутностей предметної області, інтеграція із зовнішніми джерелами, аналітика й часовий контекст, керованість і доступ. Для деталізації вимоги керування доступом (FR-11) розроблено матрицю прав за моделлю CRUD,

наведену в Додатку А, що регламентує повноваження чотирьох базових ролей — науковця, адміністратора підрозділу або аналітика, системного адміністратора та зовнішнього API.

За четвертим завданням сформовано специфікацію нефункціональних вимог, що містить 9 вимог (NFR-01...NFR-09, таблиця 2.2), специфікованих за моделлю якості програмних продуктів ISO/IEC 25010, моделлю якості даних ISO/IEC 25012 та рекомендаціями OWASP Cheat Sheet Series. Визначено цільові показники для кожної вимоги, зокрема час відгуку інтерфейсу до 2 секунд для типових операцій і до 5 секунд для агрегованих аналітичних запитів, підтримка зростання обсягу записів до 10^6 , розмежування доступу між чотирма ролями користувачів, обов'язкова фіксація статусу верифікації для кожного значення показника та сумісність зі стандартними ідентифікаторами ORCID, DOI, ROR, ISSN.

За п'ятим завданням визначено основні сутності предметної області — науковець, профіль науковця, публікація, джерело публікації, наукометрична платформа, наукометричний показник, значення показника, афіліація, структурний підрозділ, зовнішній ідентифікатор, джерело даних і часовий зріз. На їх основі побудовано концептуальну ER-модель і логічну реляційну модель бази даних, що включає таблиці Scientists, ScientistProfiles, Publications, PublicationAuthors, Sources, Platforms, MetricTypes, MetricValues, Units, Affiliations, ExternalIdentifiers, DataSources і Snapshots із системою первинних і зовнішніх ключів та обмеженнями цілісності. Для забезпечення інтеграції із зовнішніми джерелами розроблено схему мапінгу JSON-метаданих Crossref REST API у структуру реляційної моделі (Додаток Б), а практичну реалізацію центральних таблиць подано у вигляді DDL-скрипту мовою SQL (Додаток В).

За шостим завданням обґрунтовано узагальнений підхід до представлення наукометричних показників у структурі даних системи.

Доведено, що для розвитку `publication.kspu.edu` недоцільно зберігати кожен нову метрику як окреме ізольоване поле. Запропоновано модель, у якій показники описуються через універсальний набір критеріїв, об'єднаних у три логічні групи (таблиця 3.2): критерії ідентифікації показника (назва, тип, формат), критерії контексту вимірювання (об'єкт застосування, платформа-джерело, часовий зріз) і критерії метаданих запису (джерело даних, статус верифікації). Такий підхід реалізований через довідник типів показників `MetricTypes` і узагальнену таблицю значень `MetricValues`, що забезпечує масштабованість моделі: додавання нової платформи або нової метрики не потребує зміни структури центральних таблиць, а зводиться до поповнення довідників. Сформований перспективний перелік платформ і метрик включає Scopus (h-індекс, CiteScore, SNIP, SJR), Web of Science (профілі дослідників, JCR), Google Scholar (h-індекс, i10-індекс, h5-показники), ORCID (стійка ідентифікація), Crossref та OpenAlex (відкриті метадані).

Практичне значення одержаних результатів полягає в тому, що сформована специфікація вимог (FR-01...FR-11 та NFR-01...NFR-09), розроблені концептуальна і логічна моделі даних, узагальнений підхід до представлення показників і прикладні артефакти (CRUD-матриця, схема мапінгу метаданих, DDL-скрипт центральних таблиць) можуть бути використані як основа для подальшого проектування та вдосконалення `publication.kspu.edu` в межах цифрової інфраструктури Херсонського державного університету. Запропоновані підходи створюють передумови для впорядкованого та формалізованого накопичення даних, розвитку багатоплатформної інтеграції, підтримки історичних зрізів і розширення аналітичних сервісів — важливих для внутрішнього моніторингу, рейтингування та прийняття управлінських рішень у закладі вищої освіти. Перспективним напрямом подальших досліджень є реалізація запропонованої моделі у вигляді функціонального бекенд-застосунку та інтеграція до архітектури

системи елементів машинного навчання для автоматизованої верифікації профілів і нормалізації метаданих.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Hood W. W., Wilson C. S. The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*. 2001. Vol. 52, No. 2. P. 291–314.
2. Broadus R. N. Toward a definition of «bibliometrics». *Scientometrics*. 1987. Vol. 12, No. 5–6. P. 373–379.
3. Рейтинг. publication.kspu.edu : вебсайт. URL: <https://publication.kspu.edu/> (дата звернення: 15.03.2026).
4. Новини сервісу publication.kspu.edu. publication.kspu.edu : вебсайт. URL: <https://publication.kspu.edu/post/> (дата звернення: 15.03.2026).
5. Аналітика. publication.kspu.edu : вебсайт. URL: <https://publication.kspu.edu/analytics/> (дата звернення: 15.03.2026).
6. Рейтинг науково-педагогічних працівників. publication.kspu.edu : вебсайт. URL: <https://publication.kspu.edu/rating/persons/> (дата звернення: 15.03.2026).
7. Положення про кваліфікаційну роботу (проект). Херсонський державний університет : вебсайт. URL: <https://www.kspu.edu/About/Faculty/FBP/ChairBranchLaw/qualificationworks.aspx> (дата звернення: 15.03.2026).
8. Про затвердження стандарту вищої освіти за спеціальністю 126 «Інформаційні системи та технології» для першого (бакалаврського) рівня вищої освіти : наказ Міністерства освіти і науки України від 12.12.2018 № 1380. URL: <https://mon.gov.ua/static-objects/mon/uploads/public/661/685/f2c/661685f2cdfa1070210940.pdf> (дата звернення: 15.03.2026).
9. Yang S., Yuan Q. Are Scientometrics, Informetrics, and Bibliometrics Different? *Proceedings of the 16th International Conference on Scientometrics & Informetrics (ISSI 2017)*. Wuhan, China, 2017. P. 1507–1518.

10. Price D. J. de Solla. *Little Science, Big Science*. New York : Columbia University Press, 1963. 119 p.
11. Garfield E. Citation indexes for science: a new dimension in documentation through association of ideas. *Science*. 1955. Vol. 122, No. 3159. P. 108–111. DOI: <https://doi.org/10.1126/science.122.3159.108>.
12. Granovsky Y. V. Is it possible to measure science? V. V. Nalimov's research in scientometrics. *Scientometrics*. 2001. Vol. 52, No. 2. P. 127–150.
13. About ORCID. ORCID : вебсайт. URL: <https://info.orcid.org/what-is-orcid/> (дата звернення: 15.03.2026).
14. Hirsch J. E. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*. 2005. Vol. 102, No. 46. P. 16569–16572. DOI: <https://doi.org/10.1073/pnas.0507655102>.
15. Hicks D., Wouters P., Waltman L., de Rijcke S., Rafols I. Bibliometrics: The Leiden Manifesto for research metrics. *Nature*. 2015. Vol. 520, No. 7548. P. 429–431.
16. San Francisco Declaration on Research Assessment. DORA : вебсайт. URL: <https://sfdora.org/read/> (дата звернення: 15.03.2026).
17. Scopus metrics. Elsevier : вебсайт. URL: <https://www.elsevier.com/products/scopus/metrics> (дата звернення: 15.03.2026).
18. CiteScore metrics. Elsevier : вебсайт. URL: <https://www.elsevier.com/products/scopus/metrics/citescore> (дата звернення: 15.03.2026).
19. Measuring a journal's impact. Elsevier : вебсайт. URL: <https://www.elsevier.com/researcher/author/tools-and-resources/measuring-a-journals-impact> (дата звернення: 15.03.2026).
20. Web of Science platform. Clarivate : вебсайт. URL: <https://clarivate.com/academia-government/scientific-and-academic->

- research/research-discovery-and-referencing/web-of-science/ (дата звернення: 15.03.2026).
21. Web of Science Researcher Profiles. Clarivate : вебсайт. URL: <https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/wos-researcher-profiles/> (дата звернення: 15.03.2026).
22. Journal Citation Reports. Clarivate : вебсайт. URL: <https://clarivate.com/academia-government/scientific-and-academic-research/research-funding-analytics/journal-citation-reports/> (дата звернення: 15.03.2026).
23. Google Scholar Citations. Google Scholar : вебсайт. URL: <https://scholar.google.com/citations> (дата звернення: 15.03.2026).
24. Google Scholar Metrics. Google Scholar : вебсайт. URL: https://scholar.google.com/citations?view_op=metrics_intro (дата звернення: 15.03.2026).
25. ORCID for Researchers. ORCID : вебсайт. URL: <https://info.orcid.org/researchers/> (дата звернення: 15.03.2026).
26. Metadata Retrieval and REST API. Crossref : вебсайт. URL: <https://www.crossref.org/documentation/retrieve-metadata/rest-api/> (дата звернення: 15.03.2026).
27. Participation Reports. Crossref : вебсайт. URL: <https://www.crossref.org/documentation/reports/participation-reports/> (дата звернення: 15.03.2026).
28. Sources overview. OpenAlex technical documentation : вебсайт. URL: <https://docs.openalex.org/api-entities/sources> (дата звернення: 15.03.2026).
29. Dimensions platform. Dimensions : вебсайт. URL: <https://www.dimensions.ai/> (дата звернення: 15.03.2026).

30. Dimensions Linked Research Data. Dimensions : вебсайт. URL:
<https://www.dimensions.ai/resources/dimensions-linked-research-data/>
 (дата звернення: 15.03.2026).
31. Article Metrics in Dimensions. Dimensions: вебсайт. URL:
<https://www.dimensions.ai/resources/article-metrics-in-dimensions/> (дата
 звернення: 15.03.2026).
32. Scopus Author Profiles. Elsevier: вебсайт. URL:
<https://www.elsevier.com/products/scopus/author-profiles> (дата
 звернення: 15.03.2026).
33. Authors overview. OpenAlex technical documentation: вебсайт. URL:
<https://docs.openalex.org/api-entities/authors> (дата звернення:
 15.03.2026).
34. What should I do if an author has multiple author IDs? Elsevier Support :
 вебсайт. URL:
https://service.elsevier.com/app/answers/detail/a_id/28238/ (дата
 звернення: 15.03.2026).
35. Insights from a roundtable on author affiliation metadata. Crossref blog :
 вебсайт. URL: [https://www.crossref.org/blog/insights-from-a-roundtable-
 on-author-affiliation-metadata/](https://www.crossref.org/blog/insights-from-a-roundtable-on-author-affiliation-metadata/) (дата звернення: 15.03.2026).
36. Електронний сервіс рейтингування Publication. Херсонський
 державний університет : вебсайт. URL:
[https://www.kspu.edu/About/DepartmentAndServices/Library/Actual/Sci
 entometricActivity/ElectronicRatingServicePublication.aspx](https://www.kspu.edu/About/DepartmentAndServices/Library/Actual/ScientometricActivity/ElectronicRatingServicePublication.aspx) (дата
 звернення: 15.03.2026).
37. Про запровадження в університеті електронного сервісу
 рейтингування : наказ Херсонського державного університету від
 23.04.2020 № 385-Д. Херсонський державний університет : вебсайт.
 URL:
https://www.kspu.edu/FileDownload.ashx/Наказ%20від%2023.04.2020_

285-Д.pdf?id=28e14496-8463-448c-9246-bb1c27ea5bc4 (дата звернення: 15.03.2026).

38.Рейтинг факультетів. publication.kspu.edu : вебсайт. URL: <https://publication.kspu.edu/rating/faculties/> (дата звернення: 15.03.2026).

39.Рейтинг кафедр. publication.kspu.edu : вебсайт. URL: <https://publication.kspu.edu/rating/departments/> (дата звернення: 15.03.2026).

40.Документи за галузями знань. publication.kspu.edu : вебсайт. URL: <https://publication.kspu.edu/doc-knowledge/> (дата звернення: 15.03.2026).

41.Наукометрична діяльність. Херсонський державний університет : вебсайт. URL: <https://www.kspu.edu/About/DepartmentAndServices/Library/Actual/ScientometricActivity.aspx> (дата звернення: 15.03.2026).

42.Звіт роботи Наукової бібліотеки ХДУ за 2022–2023 н. р. Херсонський державний університет : вебсайт. URL: <https://www.kspu.edu/FileDownload.ashx/ЗВІТ%20РОБОТИ%20Наукової%20бібліотеки%20ХДУ%20за%202022-2023%20н.р.pdf?id=c83d948d-9e4f-4dcd-83f9-8303460b493d> (дата звернення: 15.03.2026).

43.ISO/IEC 25010:2011. Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. ISO : вебсайт. URL: <https://www.iso.org/standard/35733.html> (дата звернення: 15.03.2026).

44.ISO/IEC 25010:2011 (online). Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. ISO Online Browsing Platform : вебсайт. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en> (дата звернення: 15.03.2026).

45. Authentication Cheat Sheet. OWASP Cheat Sheet Series : вебсайт. URL:
https://cheatsheetseries.owasp.org/cheatsheets/Authentication_Cheat_Sheet.html (дата звернення: 15.03.2026).
46. Authorization Cheat Sheet. OWASP Cheat Sheet Series : вебсайт. URL:
https://cheatsheetseries.owasp.org/cheatsheets/Authorization_Cheat_Sheet.html (дата звернення: 15.03.2026).
47. OAuth 2.0 Protocol Cheatsheet. OWASP Cheat Sheet Series : вебсайт. URL:
https://cheatsheetseries.owasp.org/cheatsheets/OAuth2_Cheat_Sheet.html (дата звернення: 15.03.2026).
48. REST Security Cheat Sheet. OWASP Cheat Sheet Series : вебсайт. URL:
https://cheatsheetseries.owasp.org/cheatsheets/REST_Security_Cheat_Sheet.html (дата звернення: 15.03.2026).
49. Session Management Cheat Sheet. OWASP Cheat Sheet Series : вебсайт. URL:
https://cheatsheetseries.owasp.org/cheatsheets/Session_Management_Cheat_Sheet.html (дата звернення: 15.03.2026).
50. ISO/IEC 25012:2008. Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model. ISO : вебсайт. URL: <https://www.iso.org/standard/35736.html> (дата звернення: 15.03.2026).
51. Overview. OpenAlex technical documentation : вебсайт. URL:
<https://docs.openalex.org/> (дата звернення: 15.03.2026).
52. Works overview. OpenAlex technical documentation : вебсайт. URL:
<https://docs.openalex.org/api-entities/works> (дата звернення: 15.03.2026).
53. Documentation. ORCID: вебсайт. URL:
<https://info.orcid.org/documentation/> (дата звернення: 15.03.2026).

54. Institutions Overview. OpenAlex Developers : вебсайт. URL: <https://developers.openalex.org/api-reference/institutions> (дата звернення: 15.03.2026).
55. Pure: Comprehensive Research Information Management. Elsevier : вебсайт. URL: <https://www.elsevier.com/products/pure> (дата звернення: 15.03.2026).
56. DSpace-CRIS Home. LYRISIS Wiki : вебсайт. URL: <https://wiki.lyrasis.org/display/DSPACECRIS> (дата звернення: 15.03.2026).
57. VIVO Project: Connect, Share, Discover. DuraSpace: вебсайт. URL: <https://duraspace.org/vivo/> (дата звернення: 15.03.2026).

ДОДАТКИ

Додаток А

Матриця прав доступу користувачів до інформаційної системи
(модель CRUD).

Таблиця А.1

Матриця прав доступу (Create, Read, Update, Delete)

Сутність предметної області	Науковець (Власник профілю)	Адміністратор підрозділу / Аналітик	Системний адміністратор	Зовнішній API (Інтеграція)
Власний профіль	Read, Update	Read	Read, Update, Delete	Read
Профілі інших працівників	Немає доступу	Read	Read, Update, Delete	Read
Публікації (ручне введення)	Create, Read, Update	Read	Create, Read, Update, Delete	Немає доступу
Імпортвані дані (API)	Read (запит на уточнення)	Read	Read, Delete (скасування)	Create, Update
Довідник підрозділів (Units)	Read	Read	Create, Read, Update, Delete	Немає доступу

Аналітич ні дашборди / Рейтинги	Read (свої та загальні)	Read (свій підрозділ)	Read (всі рівні)	Немає доступу
Зовнішні ідентифіка тори	Create, Read	Read	Read, Update, Delete	Read, Update

Додаток Б

Схема мапінгу метаданих від Crossref REST API у реляційну базу даних.

Таблиця Б.1

Мапінг атрибутів JSON-відповіді у структуру таблиць бази даних

Поле у JSON-відповіді Crossref API	Цільова таблиця / поле	Тип обробки / валідація
message.items[].DOI	ExternalIdentifiers.identifier_value	Збереження DOI як зовнішнього ідентифікатора публікації; перевірка на дублювання
message.items[].title[0]	Publications.title	Прямий запис із очищенням від службових тегів
message.items[].type	Publications.document_type	Нормалізація типу документа (journal-article, book-chapter тощо)
message.items[].published.date-parts[0][0]	Publications.year	Витягнення року з масиву дат
message.items[].container-title[0]	Sources.source_name	Пошук або створення запису в таблиці Sources; використовується разом з іншими атрибутами джерела
message.items[].ISSN[]	ExternalIdentifiers.identifier_value / Sources	Використання ISSN як надійнішого ідентифікатора джерела публікації
message.items[].author[].ORCID	ExternalIdentifiers.identifier_value	Пошук або створення зовнішнього ідентифікатора автора; подальше зіставлення з внутрішнім записом науковця
message.items[].author[].given + message.items[].author[].family	Scientists / PublicationAuthors	Використання для зіставлення автора за ПІБ, якщо ORCID відсутній
message.items[].URL	Publications.url або службове поле джерела	Збереження посилання на запис Crossref
message.items[].publisher	Sources.publisher	Прямий запис або оновлення відомостей про джерело
message.items[].license[]	Publications / службове поле метаданих	За потреби збереження інформації про ліцензію
message.items[].funder[]	Publications /	Опціональне розширення

Поле у JSON-відповіді Crossref API	Цільова таблиця / поле	Тип обробки / валідація
	окрема таблиця фінансування	моделі в разі підтримки фінансування
дата імпорту запису	Snapshots.snapshot_date	Фіксація дати отримання метаданих як часового зрізу
спосіб надходження запису	DataSources.source_type	Значення API для записів, отриманих через Crossref REST API
статус перевірки запису	MetricValues.verification_status / службовий статус публікації	Початкове значення Imported з можливістю подальшої зміни

Під час імпорту метаданих від Crossref REST API DOI доцільно використовувати як основний зовнішній ідентифікатор публікації, а ORCID — як зовнішній ідентифікатор автора, що потребує подальшого зіставлення з внутрішнім записом науковця. Поле `container-title` може використовуватися для первинного визначення джерела публікації, однак для надійнішого зв'язування доцільно враховувати також ISSN або інші стабільні ідентифікатори джерела. Такий підхід забезпечує коректніше перетворення JSON-метаданих у структуру реляційної моделі та зменшує ризик дублювання або хибного зіставлення записів [26; 35].

Додаток В

Фрагмент SQL-скрипту створення центральних таблиць бази даних (DDL).

-- Створення довідника типів показників

```
CREATE TABLE MetricTypes (  
    metric_type_id SERIAL PRIMARY KEY,  
    metric_name VARCHAR(150) NOT NULL,  
    metric_description VARCHAR(255),  
    application_level VARCHAR(50) NOT NULL,  
    metric_kind VARCHAR(50) NOT NULL,  
    value_format VARCHAR(50) NOT NULL,  
    unit_name VARCHAR(50),  
    metric_group VARCHAR(100),  
    is_active BOOLEAN DEFAULT TRUE  
);
```

-- Створення довідника наукометричних платформ

```
CREATE TABLE Platforms (  
    platform_id SERIAL PRIMARY KEY,  
    platform_name VARCHAR(100) NOT NULL,  
    has_api BOOLEAN DEFAULT FALSE,  
    base_url VARCHAR(255)  
);
```

-- Створення довідника джерел даних (походження записів)

```
CREATE TABLE DataSources (  
    data_source_id SERIAL PRIMARY KEY,  
    source_type VARCHAR(20) NOT NULL,  
    source_description VARCHAR(255),
```

```

CONSTRAINT chk_data_source_type
    CHECK (source_type IN ('API', 'Manual', 'Import'))
);

-- Створення таблиці часових зрізів
CREATE TABLE Snapshots (
    snapshot_id SERIAL PRIMARY KEY,
    snapshot_date DATE NOT NULL,
    snapshot_type VARCHAR(50),
    snapshot_description VARCHAR(255)
);

-- Створення центральної поліморфної таблиці значень показників
CREATE TABLE MetricValues (
    metric_value_id SERIAL PRIMARY KEY,
    metric_type_id      INT      NOT      NULL      REFERENCES
MetricTypes(metric_type_id),
    platform_id        INT      NOT      NULL      REFERENCES
Platforms(platform_id),
    data_source_id     INT      NOT      NULL      REFERENCES
DataSources(data_source_id),
    snapshot_id        INT      NOT      NULL      REFERENCES
Snapshots(snapshot_id),
    numeric_value NUMERIC(10,2) NOT NULL,
    object_type VARCHAR(50) NOT NULL,
    object_id INT NOT NULL,
    verification_status  VARCHAR(20)  NOT  NULL  DEFAULT
'Imported',
    CONSTRAINT chk_object_type

```

```
        CHECK (object_type IN ('Scientist', 'Publication', 'Source',  
'Unit')),  
        CONSTRAINT chk_verification_status  
        CHECK (verification_status IN ('Imported', 'Confirmed', 'Manual',  
'Rejected'))  
    );
```

-- Створення індексів для пришвидшення аналітичних та
рейтингових запитів

```
    CREATE INDEX idx_metric_object ON MetricValues (object_type,  
object_id);  
    CREATE INDEX idx_metric_platform ON MetricValues  
(platform_id);  
    CREATE INDEX idx_metric_type ON MetricValues (metric_type_id);  
    CREATE INDEX idx_metric_snapshot ON MetricValues  
(snapshot_id);
```