УДК 004.65
DOI: https://doi.org/10.11603/mie.1996-1960.2022.1-2.13104

НОВІ ТЕНДЕНЦІЇ РОЗВИТКУ СИСТЕМ ПРЕДСТАВЛЕННЯ ТА УПРАВЛІННЯ ДАНИМИ. АНАЛІТИЧНИЙ ПОГЛЯД

О. П. Мінцер, Л. Ю. Бабінцева

Національний університет охорони здоров'я України імені П. Л. Шупика

Проаналізовано проблеми сучасного розвитку систем представлення та управління даними. Кількість, різноманітність та швидкість даних становлять особливу цінність для управління, науки, виробництва. Сьогодні великі дані використовують для низки завдань: удосконалення та прискорення процесів розроблення продуктів, забезпечення високої якості роботи та надання послуг, високого рівня безпеки, операційної ефективності тощо. Зі більшенням обсягів великих даних відкриваються нові можливості. В роботі представлено можливості застосування метатехнологій і метаданих в охороні здоров'я. Особливу увагу приділено новій технології GO FAIR.

Постійне лавиноподібне зростання даних обумовлює безперервне зростання проблем управління ними. Правильне управління даними корисне для пацієнтів, надавачів послуг і страховиків, має далекосяжні наслідки для здоров'я всього населення. Серед сучасних напрямів систем управління даними інтерес представляє платформа GO FAIR, що являє собою технологію управління без втручання користувача, по суті, один із перших кроків до формування цифрової інфраструктури.

Ключові слова: системи представлення та управління даними, великі дані, хмарні технології, туманні обчислення, озера даних, технологія GO FAIR.

NEW TRENDS IN THE DEVELOPMENT OF DATA PRESENTATION AND MANAGEMENT SYSTEMS. ANALYTICAL VIEW

O. P. Mintser, L. Yu. Babintseva

Shupyk National Healthcare University of Ukraine

Background. IThe problems of the modern development of data presentation and management systems are analyzed. The amount, variety and speed of data are of particular value for management, science, and production. Today, big data is used for a number of tasks: improvement and acceleration of product development processes, ensuring high quality of work and service provision, high level of security, operational efficiency, etc. As the volume of big data grows, new opportunities open up. Special attention is paid to the new GO FAIR technology. The purpose of the study was to present the possibilities of using metatechnologies and metadata in health care.

Materials and methods. A theoretical analysis and generalization of information on the use of metadata and metatechnologies in health care was carried out, as well as the systematization of research results on the specified topic using the following databases of scientific periodicals: Web of Science, PubMed, Scopus, ScienceDirect, OSF, ORE, IEEE, EBSCO. Classical methods of information search were applied at various stages of the research: selective, intuitive, inductive, deductive, and the method of bibliographic references. The methods of system analysis, structural analysis and design (SADT), data mining, cluster and factor analysis, system approach and methods of decision-making theory were used to process the received information.

Results. The volume of data coming from various sources (video cameras, social networks, audio recordings, IoP devices) is constantly increasing, which leads to the emergence of new big data management systems. The most important step for continuous software delivery is continuous integration. Cloud computing has advantages for global analysis on longer time scales, where latency is not a concern. Health data management is tasked not only with organizing medical data, but also with integrating and analyzing it to make patient care more efficient and to obtain information that can improve medical outcomes while protecting data privacy and security. Among modern areas of data management systems, the GO FAIR platform is considered in detail.

Conclusions. The constant avalanche-like growth of data determines the continuous growth of data management problems. Good data management benefits patients, providers and insurers, and has far-reaching implications for the health of the entire population. Among modern directions of data management systems, the GO FAIR platform is of interest, which is a management technology without user intervention, in fact, one of the first steps towards the formation of a digital infrastructure.

Keywords: data presentation and management systems, Big Data, cloud technologies, fog computing, data lakes, GO FAIR technology.

© О. П. Мінцер, Л. Ю. Бабінцева

Вступ. Управління даними передбачає ефективну, економічну та безпечну організацію процесів збирання, зберігання та використання даних. Його мета полягає в оптимізації управління даними працівників, компаній, підключеними пристроями з дотриманням політик і правил таким чином, щоб вони могли приймати рішення та діяти найвигідніше для підприємства. Надійна стратегія управління даними набуває все більшого значення зі зростанням кількості підприємств, що покладаються на нематеріальні активи для створення прибутку.

В охороні здоров'я системи представлення та управління даними мають забезпечити інформацією стосовно:

• стану здоров'я населення та впливу соціальних детермінант здоров'я. Для цього необхідна більш тісна інтеграція даних про первинну медико-санітарну допомогу, об'єднання даних приватних постачальників послуг, даних про захворюваність, психічне здоров'я, соціальні детермінанти та нерівності людей, даних, що інформують про стан здоров'я протягом усього життєвого циклу;

• якості надання медичних послуг: сюди входять дані про ефективність надання комплексної допомоги, документація про способи надання медичної допомоги, дані, що дозволяють проводити порівняльний аналіз, різні параметри результатів надання медичних послуг;

• наявних ресурсів, а саме: кадрових ресурсів, об'єктах та обладнанні, послугах у системі охорони здоров'я, потенціалі та оцінюванні ефективності роботи системи охорони здоров'я.

Управління цифровими даними включає широкий набір завдань, політик, процедур і практик. Робота з управління даними повинна враховувати безліч факторів, включаючи такі: створення різноманітних даних на будь-якому рівні, доступ до них та їх оновлення; зберігання даних у різних хмарах і локальній системі; забезпечення високої доступності та відновлення при катастрофічних збоях; використання даних у додатках, аналітичних засобах та алгоритмах; контроль конфіденційності та безпеки даних; архівація та знищення даних відповідно до графіків зберігання та нормативних вимог тощо.

Системи управління даними створюються на основі платформ управління та можуть включати бази даних, озера даних і сховища даних, системи управління великими даними, аналітичні засоби та багато іншого. Всі ці компоненти взаємодіють між собою, утворюючи єдину платформу для роботи з даними. Вона забезпечує засоби управління даними, що застосовуються в корпоративних додатках, а також інструменти аналізу та алгоритми оброблення цих даних. Хоча сучасні інструменти дозволяють автоматизувати виконання багатьох завдань управління, більшість розгортань баз даних має настільки великий розмір і складну структуру, що втручання адміністратора бази даних, як і раніше, є необхідним. Більш того, відсутнє рішення, що здатне забезпечувати ефективне уніфіковане керування різноманітними даними на єдиному рівні. Але це збільшує можливість появи помилок. Відповідно скорочення потреби в ручному управлінні даними – одна з основних цілей нових технологій управління даними, автономної бази даних.

Отже, створення нової методології управління даними стає очевидною проблемою. Одним із напрямів рішення являється застосування метатехнологій. Розглянемо такий підхід при вирішенні відносно простого завдання – розроблення презентацій. Відомо, що слайди презентації – важливий комунікативний засіб. Як наслідок, програмне забезпечення для презентацій являється важливим інструментом для структурування та представлення проблемних областей. Відповідні інструменти дозволяють створювати вражаючі презентації. Можна припустити, що на сьогодні накопичено велику кількість слайдів. Однак, незважаючи на певну зручність і функціональність, що пропонують інструменти, а також вражаючий вигляд професійно розроблених бізнес-презентацій, поточна практика далека від задовільної. Абсолютно слушно в роботі [2] підкреслюється, що час від часу використання програмного забезпечення для презентацій – особливо PowerPoint – піддається критиці за псування творчих здібностей та індивідуальних стилів презентації, використовується для управління презентаціями лише у вигляді послідовності слайдів: таким чином сприяючи нудним презентаціям, що компрометують "споглядальний аналітичний метод" [3]. Щоб протистояти цьому ефекту автор підкреслює актуальність "когнітивного підходу", який хоча і не повинен бути домінуючим через "обмеження технології презентації", але він безумовно стає більш продуктивним, цілісним і забезпечує повторне використання.

Зазвичай, презентації зберігаються у вигляді файлів. Щоразу, коли створюється нова презентація, повторне використання обмежується, що є серйозною загрозою цілісності. Більш того, продуктивність, якість і цілісність страждають від

відсутності концептуальної основи змісту. Тобто інструменти, що застосовуються для створення слайдів і вмісту, позбавлені формальної семантики. Відсутність семантики являється особливою проблемою для графічних зображень. Відсутність семантики, тобто відсутність правил, що обмежують використання предметів, хоча і сприяє деякій гнучкості, проте в кінцевому форматі забезпечує лише поверхнево представлення матеріалу. Ще більші проблеми має повторне використання даних дослідження.

Мета дослідження: представити можливості застосування метатехнологій і метаданих в охороні здоров'я.

Матеріал і методи дослідження. Проведено теоретичний аналіз і узагальнення інформації про використання в охороні здоров'я метаданих і метатехнологій, а також систематизацію результатів досліджень на зазначену тематику із використанням таких баз даних наукової періодики: Web of Science, PubMed, Scopus, ScienceDirect, OSF, ORE, IEEE, EBSCO. На різних етапах дослідження застосовано класичні методи інформаційного пошуку: вибірковий, інтуїтивний, індуктивний, дедуктивний, метод бібліографічних посилань. Для оброблення отриманої інформації використовували методи системного аналізу, структурного аналізу та проєктування (SADT), data mining, кластерний і факторний аналіз, системний підхід і методи теорії прийняття рішень.

Результати та їх обговорення. У певному сенсі термін "великі дані" (Big Data) слід розуміти буквально: він означає велику кількість даних. Проте великі дані, на відміну від традиційних, дуже різноманітні, та збираються вони швидше. Саме кількість, різноманітність і швидкість таких даних становлять особливу цінність для управління, науки, бізнесу. Компанії використовують великі дані, щоб удосконалювати та прискорювати процес розроблення продуктів, виконувати попереджувальне обслуговування, забезпечувати високу якість роботи із замовниками, а також високий рівень безпеки, операційної ефективності та низку інших переваг. Зі збільшенням обсягів великих даних відкриваються нові можливості [13-16].

Підкреслимо, що обсяг даних, який надходить із розрізнених джерел (відеокамери, соціальні мережі, аудіозаписи, пристрої Інтернету речей (IoP)) постійно збільшується, що призводить до появи нових систем управління великими даними. Аналізуючи основні сфери застосування: інтеграція великих даних, робота з даними різних типів (від пакетних до потокових) та їх перетворення для подальшого використання, зауважимо, що всі перелічені сфери не мають однозначного тлумачення та рекомендацій до їх використання. Зрозуміло, що першою проблемою можна назвати стратегію управління, що апріорно уявляється досить складною. Відповідно до дефініції, управління великими даними – це ефективне, надійне та безпечне зберігання даних в озері або сховищі даних, а також їх оброблення (часто з використанням об'єктного сховища).

Термін "озеро даних" з'явився в 2015 році, проте сама концепція застосовується вже понад десятиліття. Озеро даних можна розглядати як централізоване місце для зберігання великих обсягів необроблених даних у їхньому власному форматі. Порівняно з ієрархічним сховищем, у якому дані зберігаються в файлах і папках, в озері даних використовується плоска архітектура. За рахунок застосування тегів та ідентифікаторів для метаданих спеціалісти по роботі з великими даними можуть набагато простіше та ефективніше знаходити й отримувати дані в масштабі кількох регіонів. Можливий і аналіз великих даних – видобуток нової, корисної інформації за допомогою інструментів аналітики, включаючи аналіз графів, а також створення моделей за допомогою технологій візуалізації на базі машинного навчання та штучного інтелекту. Крім того, в озерах даних є перевага використання власного формату додатків.

Взагалі, основною системою для збирання та аналізу великих обсягів даних являється платформа управління. Комерційні платформи для управління даними, зазвичай, включають програмні інструменти для управління від постачальників систем управління базами даних (СУБД) або сторонніх вендорів. Подібні рішення для управління даними допомагають IT-спеціалістам та адміністраторам баз даних виконувати такі стандартні завдання: виявлення, діагностика та усунення помилок у системі бази даних або її інфраструктури та розсилання пов'язаних із ними повідомлень; розподіл ресурсів пам'яті та місця в базі даних; внесення змін до схеми бази даних; оптимізація оброблення запитів до бази даних для підвищення ефективності програм.

Найважливішим кроком для безперервної доставки програмного забезпечення є безшовна інтеграція (БІ). БІ – це процес, у якому додається чи інтегрується новий модуль або функція програми чи апаратного забезпечення, не спричиня-

ючи жодних помітних помилок або ускладнень і означає, що будь-яке змінення, застосоване до системи, відбувається без будь-якого негативного впливу в результаті інтеграції. Технологія безшовної інтеграції використовується в контексті розроблення програмного забезпечення та систем, а також для комп'ютерного обладнання. Практика розроблення, при якій розробники фіксують свої зміни коду (зазвичай, поступові) у централізованому репозиторії вихідного коду, що запускає набір автоматичних зборок і тестів. У цьому репозиторії розробники можуть автоматично виявляти помилки на ранньому етапі, перш ніж продукт буде передано у виробництво. Конвеєр безшовної інтеграції зазвичай включає ряд кроків, починаючи з фіксації коду та закінчуючи виконанням базової автоматизованої перевірки дотримання стандарту оформлення коду або статичного аналізу, визначення залежностей і, нарешті, складанням програмного забезпечення та виконанням тестування деяких основних модулів перед створенням артефакту складання. Системи управління вихідним кодом, такі як Github, Gitlab та інші, забезпечують інтеграцію з веб-перехоплювачами, на які інструменти БІ, такі як Jenkins, можуть підписатися для запуску автоматичної зборки та тестування після кожної реєстрації коду.

Зауважимо, що існують і автономні бази даних (Autonomous Database), які представляють собою хмарні системи з технологіями штучного інтелекту та машинного навчання для автоматизації цілого ряду завдань з управління: резервне копіювання, забезпечення безпеки та контроль ефективності, що зазвичай входять до кола завдань адміністраторів баз даних. Такі бази даних також називають самоврядними. Автономні бази даних забезпечують низку переваг: зниження складності, зниження ризику людських помилок, підвищення рівня надійності та захищеності бази даних, підвищення операційної ефективності, зниження витрат.

Резюмуючи формальну сторону постійного лавиноподібного зростання даних важливо підкреслити, що паралельно зростанню кількості даних постійно наростають проблеми управління даними, потреби в ефективних інструментах управління.

Неповний перелік найскладніших проблем, із якими сьогодні стикаються системи управління даними включає:

1. Відсутність (недостатність) цінної інформації у даних. Дані будуть марними, якщо не має інформації про їх зберігання та використання. Рішення про управління даними мають бути масштабованими та високопродуктивними, щоб можна було своєчасно отримувати відповідну аналітичну інформацію.

2. Важко підтримувати ефективність управління даними на високому рівні. Потрібна організація сучасного моніторингу запитів і відповідей на них без зниження ефективності.

3. Важко задовольняти мінливим вимогам до даних. Нормативні вимоги складні, охоплюють різні юрисдикції та постійно змінюються. Особливу увагу при цьому потрібно приділяти відповідності міжнародним вимогам про забезпечення конфіденційності даних.

4. Створення та постійне вдосконалення принципів оброблення та перетворення даних. Знайдені та зібрані дані самі по собі цінності не становлять. Якщо перетворення даних на зручний для аналізу вид займає занадто багато часу та сил, то в результаті проаналізувати нічого не вдасться.

5. Постійна необхідність ефективного збереження даних. Збереження інформації здійснюється відразу в кількох системах, включаючи сховища даних і неструктуровані озера даних, де в одному репозиторії можуть розміщуватись будь-які дані в будь-якому форматі та представляти їх у тому вигляді та моделі, що будуть підходити для аналізу практично будь-якого типу.

6. Потреба постійно оптимізувати витрати на IT і гнучкість. Завдяки доступності хмарних систем управління даними компанії тепер можуть обирати, зберігати та аналізувати дані в локальному середовищі, у хмарі чи у гібридному варіанті. IT-організаціям необхідно оцінити, наскільки ідентичні локальне та хмарне середовища, щоб підтримувати максимальну гнучкість IT і знижувати витрати.

Особливості управління медичними даними. Управлінню медичними даними присвячено багато наукових досліджень [11, 12]. Зауважимо, що управління даними про здоров'я (Health Data Management – HDM), також відоме як управління інформацією про здоров'я (Health Information Management – HIM) – це систематична організація даних про здоров'я пацієнта в цифровій формі. При чому останнє може бути чим завгодно: від електронних медичних записів (EMR), створених у результаті відвідувань лікаря, електронних записів про здоров'я пацієнтів (EHR) до документів, відсканованих із історій захворювань.

Управління даними про здоров'я має завдання не лише впорядкувати медичні дані, а й інтегрувати їх, надати можливість їх аналізу, щоб зробити супровід пацієнта більш ефективним, а також отримати інформацію, яка може покращити медичні результати, захищаючи при цьому конфіденційність і безпеку даних.

Оскільки медичні практики продовжують використовувати більш складні системи електронних медичних карт (EHR), потреба в ефективному управлінні даними про здоров'я зростає. Управління даними про здоров'я – це процес, що використовується для запису, зберігання, захисту та аналізу даних, які отримуються з різних джерел. Ефективне управління даними охорони здоров'я дозволяє медичним працівникам створити всебічне уявлення про стан пацієнта.

Ефективні методи управління даними – життєво важливі, щоб не відставати від величезного обсягу даних, які щомісяця виробляються медичними закладами. Вони також допомагають надавати більш персоналізований догляд, ефективніше спілкування та відповідність нормативним вимогам.

Правильне управління даними корисно для пацієнта, постачальників послуг і страховиків, має далекосяжні наслідки для здоров'я всього населення. Зокрема, управління даними про стан здоров'я допомагає:

– створити повне уявлення про здоров'я пацієнта за допомогою інтеграції даних, зібраних із різних джерел;

– підвищити залучення пацієнтів за допомогою прогнозного аналізу та швидшої діагностики на основі наявних даних;

– покращати відстеження проблем зі здоров'ям і передбачати появу ускладнень;

– надати дані для прийняття ефективних бізнес-рішень, що підвищать ефективність медичної установи;

– дотримуватися стандартизованих планів догляду, що покращить лікування захворювання.

Для управління даними про здоров'я пацієнтів досить ефективним виявилось застосування інструменту Enterprise Data Warehouse (EDW). Він широко використовується для збору та агрегування даних в організаціях охорони здоров'я. EDW призначений для об'єднання даних із різних джерел або систем в єдине сховище даних. Це дає змогу користувачам ретельно аналізувати дані, що існували в різних системах, таких як системи управління практикою, програмне забезпечення для управління фінансами та EHR.

Зі швидким зростанням даних, пов'язаних зі здоров'ям, включаючи геномні, протеомні, візуалізаційні та клінічні, важке завдання інтеграції даних може бути перевантажено складністю середовища, включаючи розмір і різноманітність даних. Відповідно зростає роль стратегій інтеграції даних для прогнозної аналітики великих даних у дослідженнях точної медицини. Набуває значення методологія інфраструктури як коду та засобу інтеграції та управління даними. Виникає питання про створення трансляційних технологій для надання можливості організаціям охорони здоров'я інтегрувати та використовувати інфраструктуру для прискорення впровадження точної медицини.

Інтеграція даних об'єднує дані з різних джерел в єдиний набір для покращення аналітики. Наприклад, методи інтеграції охорони здоров'я можуть збирати дані з різних систем електронних медичних карт в одне всеосяжне джерело, надаючи лікарям усю необхідну інформацію для точного, стандартизованого лікування.

Обсяг медичних даних, що очікують на використання, величезний. Під час кожного відвідування лікарі та інші постачальники послуг збирають особисту інформацію та історію хвороби пацієнта, що вони можуть використовувати для визначення ефективних профілактичних стратегій, усунення неефективності та економії витрат. Погана інтеграція даних може призвести до неповної та фрагментованої інформації про пацієнта, що в кінцевому підсумку призведе до скомпрометованої допомоги та втрати критичних ресурсів.

Іншу загрозу несуть неструктуровані дані, що становлять серйозну проблему для медичних закладів, незважаючи на використання систем електронних медичних записів. Проте можливість застосування інструментів інтеграції даних здатна зробити свої дані більш значущими.

Комп'ютерне управління даними. Модель "Інфраструктура як код (IaC)", що іноді називають "програмованою інфраструктурою", – це модель, за якою процес налаштування інфраструктури аналогічний процесу програмування. По суті, дана модель являється основою хмарних обчислень, а також нею започатковано усунення кордонів між написанням додатків і створенням середовищ для цих додатків. Програми можуть містити скрипти, що створюють власні віртуальні машини та управляють ними. Отже, інфраструктура як код – це метод формального визначення закодованих інструкцій про те, як слід управляти та забезпечити

набір комп'ютерів. Інструкції виконуються таким чином, що інфраструктура, передбачена кодом, реалізується на одному або кількох комп'ютерах. Можливість обробляти інфраструктуру як код і використовувати ті ж інструменти, що й будь-який інший програмний проєкт, дозволяє розробникам швидко розгортати програми.

Існує кілька додатків, що можна застосовувати як частину стратегії інфраструктури як коду для кодування та виконання інтеграції даних, прогнозного аналізу для точної медицини. Зупинимось на деяких із них. Ключовий компонент підходу включає віртуалізацію, в якій попередньо визначено комп'ютер, що кодується як віртуальна машина, а програма, яка називається гіпервізором на цільовому комп'ютері, може використовуватися для запуску віртуальної машини. Гіпервізор можна розглядати як "гіперсупервізор", що управляє кожною дією, яку виконує віртуальна машина в цільовому середовищі. Гіпервізор захищає віртуальну машину від проблем із загальною системою управління цільовим середовищем, дозволяючи їй виконувати свої процеси. Інфраструктура, по суті, являється інкапсульованим набором процесів, що можна створити та протестувати локально, а потім розповсюдити у вигляді клонів іншим користувачам/сайтам для повторення аналізу. Це особливо актуально для великих даних, де більш рентабельно проводити аналізи там, де дані знаходяться.

Методологія "Інфраструктура як код" підтримує прийняття рішень під час різних ситуацій із комбінацією навмисних і непередбачуваних стратегій, у яких віртуалізована інфраструктура може бути повністю спланована та закодована для досягнення бажаної мети або може з'явитися шляхом адаптивної та гнучкої розробки в динамічній взаємодії із середовищем. Розгортання інфраструктури може відбуватися стратегічно, використовуючи правильні ресурси для конкретного проєкту, підтримуючи внутрішні обчислення в тумані, а також під час хмарних обчислень. Хмарні обчислення мають переваги для глобального аналізу на більш тривалих часових масштабах, коли затримка передавання даних не викликає занепокоєння.

Серед сучасних напрямів систем управління даними розглянемо детально платформу GO FAIR, що представляє технологію управління без втручання користувача. По суті, вона – один із перших кроків до формування цифрової інфраструктури, що забезпечує трансфер наукових результатів у форму, зрозумілу інвесторам, чиновникам, суспільству та забезпечує контроль за обсягами наукових даних.

Технологія GO FAIR. Історія виникнення технології досить коротка. У березні 2016 року, трохи більше ніж через 2 роки після семінару Центру Лоренца, що ввів у обіг широко відому абревіатуру, в журналі Scientific з'явився короткий коментар.

У 2016 році Європейська комісія запустила Європейську відкриту наукову хмару (The European Open Science Cloud – EOSC), що вже централізовано була заснована на Керівних принципах FAIR. Самі принципи FAIR складаються з 15 розділів. Група експертів високого рівня EOSC рекомендувала реалізувати цю інфраструктуру як середовище для спільного та повторного використання наукових даних. Оскільки вона базувалася на існуючих системо утворюючих елементах у державах-членах, мала полегшене міжнародне керівництво та управління, EOSC отримала великий ступінь свободи відносно практичної реалізації. Рекомендації EOSC стали відправною точкою для ініціативи GO FAIR.

Перші країни держав-членів ЄС, у тому числі Нідерланди, запустили ініціативу GO FAIR на початку 2018 року. Ця ініціатива є самоврядною та очолюваною спільнотою ініціативою, що працює у різних дисциплінах і країнах, спрямована на реалізацію FAIR.

GO FAIR пропонує відкриту та інклюзивну екосистему для окремих осіб, установ та організацій, що працюють разом через мережі реалізації. Мережі впровадження є основними рушійними силами GO FAIR як відкриті, інклюзивні, керовані спільнотою та самокеровані консорціуми, що працюють у різних дисциплінах та країнах. Окремі особи, установи та організації з усього світу беруть участь у різноманітних мережах впровадження GO FAIR.

GO FAIR має міжнародний офіс підтримки, координації для підтримки впровадження та подальшого розвитку (GFISCO). Він розташований у трьох місцях (Париж, Гамбург, Лейден) і фінансується міністерствами науки Франції, Німеччини та Нідерландів. Ініціатива містить базові принципи покращення можливостей пошуку, забезпечення доступу до даних, їх сумісності та, що особливо важливо, повторного використання. В 2016 році було представлено "Керівний принцип FAIR управління науковими даними" [8].

Відповідно до FAIR, функції пошуку, вилучення та подання даних реалізують не користувачі, а інформаційна система. При цьому йдеться не лише

про самі дані та метадані, а й про алгоритми та інструменти управління ними. Крім того, до розроблення підходів управління науковими даними залучаються всі зацікавлені сторони: науково-дослідні організації та окремі вчені; оператори баз даних і видання, що публікують наукові статті, результати експериментів; фінансуючи організації; виробники програмного забезпечення та інструментів оброблення даних; компанії, що надають послуги з аналізу та інтерпретації даних. Важливо, що до кола зацікавлених сторін також включаються самі обчислювальні системи (алгоритми оброблення даних) як самостійний об'єкт – залежно від їхнього рейтингу приймається рішення про включення чисельного методу конфігурації.

Сьогодні безліч європейських дослідницьких інфраструктур (DTU Library, International Neuroinformatics Coordinating Facility, Radboud University тощо) використовують концепцію FAIR при наданні доступу до своїх наукових даних. Створено і розвиваються методичні рекомендації та інструкції про подання даних відповідно до FAIR [6, 7, 9, 10]. Встановлено, що використання підходу FAIR до управління даними, дозволяє ЄС заощадити понад 10 млрд. євро на рік.

У рамках програми "Горизонт-2020" ініційовано проєкт PaNOSC, що об'єднує шість великих європейських дослідницьких інфраструктур (ESRF, European XFEL, CERIC-ERIC, ELI Delivery Consortium, ESS, ILL) для розвитку Європейської хмари відкритої науки (European Open Science Cloud-EO) – універсального міждисциплінарного репозиторію наукових даних із відкритим доступом для дослідників у всіх галузях. У рамках PaNOSC ученим із таких областей, як хімія, біологія, матеріалознавство тощо, надаються сервіси та інструменти для зберігання, пошуку й аналізу даних, отриманих на нейтронних і фотонних дослідницьких інфраструктурах. За рахунок використання постійних унікальних ідентифікаторів реалізується можливість передавання метаданих між сервісами, що дозволяє збільшити можливість повторного використання результатів наукового дослідження в масштабах не пов'язаних безпосередньо тематичних областей наукових досліджень.

У перспективі йдеться про забезпечення для всієї світової наукової спільноти, незалежно від тематичної галузі досліджень, доступу через EOSC до експериментальних даних від 15 нейтронних і фотонних європейських дослідницьких інфраструктур [5]. Фактично сьогодні йдеться не так про об'єкт цифрової наукової інфраструктури (база даних), як про послугу управління великими даними: реалізується механізм управління безліччю даних, доступних широкому спектру користувачів – від наукової спільноти до державних структур (грантодавців, спонсорів). У його основу покладено просування конкретних інфраструктурних проєктів, надання механізмів доступу до дослідницьких інфраструктур для міжнародного використання, і навіть обмін знаннями (протокол TASK 86-8). У сфері правового регулювання та організації механізмів взаємодії учасників інноваційного процесу Євросоюз просунувся далеко вперед, тому для України європейський досвід може стати орієнтиром у формуванні власних правил і регуляторних механізмів. Принципи FAIR можуть бути використані для кращої інтеграції дослідницьких інфраструктур із європейськими, а також як основа для розроблення аналогічних правил управління даними.

Розглядаючи керівні принципи GO FAIR, зауважимо, що для підтримки пошукових функцій інформаційному блоку (даним і метаданим) присвоюється унікальний постійний глобальний ідентифікатор, а самі дані описуються ідентифікатором, розширеним безліччю метаданих. Дані (метадані) реєструються або індексуються в доступному для пошуку ресурсі.

Для оптимізації доступу до даних пропонується керуватися такими принципами [10]: дані (метадані) можуть бути отримані за їх ідентифікатором за стандартизованими протоколами зв'язку; протокол доступу до даних відкритий і допускає використання уніфікованого протоколу доступу – за необхідності для доступу до даних можлива процедура автентифікації та авторизації, а метадані можуть бути доступні навіть за відсутності доступу до даних.

Повинна бути забезпечена сумісність даних не тільки з іншими даними, але й з додатками та інструментами їх аналізу, зберігання та оброблення: дані (метадані) використовують формальну, доступну та широко застосовувану мову опису даних; дані (метадані) використовують словники.

Легко зрозуміти, чому ці принципи повинні застосовуватися до наборів даних: основна мета – дані можна оптимально використати для аналізу та відкриття наукових знань. Але текст опублікованих статей також сповнений тверджень і, в свою чергу, є "даними". Вони повинні бути визнані та розглядатися як такі, і до цих "стверджуючих даних"

повинні застосовуватися принципи справедливості для подальшого використання.

Висновки. 1. Постійне лавиноподібне зростання даних обумовлює безперервне зростання проблем управлінням даними.

2. Серед сучасних напрямів систем управління даними інтерес представляє платформа GO FAIR – технологія управління без втручання користувача, по суті, один із перших кроків до формування цифрової інфраструктури.

Література.

1. Towards a Reference Ontology for Business Models / Andersson B., Bergholtz M., Edirisuriya A. et al. // Proceedings of the 25th International Conference on Conceptual Modeling (ER). – 2006. – P. 482-496.

2. A Conception of a Presentation Development and Management System Featuring 'Smart Slides'/ Frank U. // Enterprise Modelling and Information Systems Architectures. – 2011. – Vol. 6, № 1.

3. The Cognitive Style of PowerPoint: Picking Out Corrupts Within, 2nd ed. / Tufte E. – Graphics Press: Chesire/CT, 2006.

4. Экзабайтное хранилище научных данных / Григорьева М., Голосова М., Рябинкин Е., Климентов А. // Открытые системы. СУБД. – 2015. – № 4. – С. 14-17.

5. Recommendations on FAIR metrics for EOSC / Aronsen J. M., Hooft R. W. W., Slavec A. et al. – Publications Office of the European Union, 2021. – Режим доступу: http://op.europa.eu/en/publication-detail/-/publication/ced147c9- 53c0-11eb-b59f-01aa75ed71a1/.

6. The Quest for Research Information / Blümel I., Dietze S., Heller L. et al. // Procedia Computer Science. – 2014. – № 33. – P. 253-260.

7. Turning FAIR into reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data / Collins S., Genova F., Harrower N. et al. – Publications Office of the European Union, 2018. – Режим доступу: https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf.

8. Towards FAIR principles for research software / Lamprecht A.-L., Garcia L., Kuzak M. et al. // Data Science. – 2020. – № 3 (1). – P. 37-59.

9. RDA FAIR Data Maturity Model. FAIR Data Maturity Model: specification and guidelines. – Research Data Alliance, 2020.

10. The FAIR Guiding Principles for scientific data management and stewardship / Wilkinson M. D., Dumontier M., Aalbersberg Ij. J. et al. // Scientific Data. – 2016. – № 3 (1). – P. 160018.

11. Promoting better integration of health information systems: best practices and challenges / Michelsen K., Brand H., Achterberg P. W., Wilkinson J. R. – Copenhagen: WHO Regional Office for Europe, 2015 (Health Evidence Network (HEN) synthesis report).

12. Data integration strategies for predictive analytics in precision medicine / Frey L. J. // Per Med. – 2018. – № 15 (6). – P. 543-551.

13. Issues, Challenges and Solutions of Big Data in Information Management: An Overview / Saleh S. H., Ismail R., Ibrahim Z., Hussin N. // International Journal of Academic Research in Business and Social Sciences. – 2018. – № 8 (12). – P. 1382-1393.

14. Challenges and solutions in big data management - an overview, presented at 3rd. / Kanchi S., Sandilya S., Ramkrishna S. et al. // International Conference on Future Internet of Things and Cloud. Rome, Italy, 2015.

15. Big data analytics: challenges and applications for social media data / Krishnaveni C., Udhayakumar U. // National Conference on Emerging Trends in Computing Technologies. Villupuram, Tamil Nadu, 2018: International Research Group.

16. 3 massive big data problems everyone should know about / Marr B. – Режим доступу: https://www.forbes.com/sites/bernardmarr/2017/06/15/3-massive-big-data-problemseveryone-should-know-about/#7df0e5606186.

References.

1. Andersson, B., Bergholtz, M., Edirisuriya, A. et al. (2006). Towards a Reference Ontology for Business Models. In: Proceedings of the 25th International Conference on Conceptual Modeling (ER), 482-496.

2. Frank, U. (2011). A Conception of a Presentation Development and Management System Featuring 'Smart Slides'. Enterprise Modelling and Information Systems Architectures. Vol. 6, No. 1.

3. Tufte, E. (2006). The Cognitive Style of PowerPoint: Picking Out Corrupts Within, 2nd ed. Graphics Press: Chesire/CT.

4. Grigoryeva, M., Golosova, M., Ryabinkin, E., Klimentov, A. (2015). Exabyte repository of scientific data [Ekzabaytnoye khranilishche nauchnykh dannykh]. Open Systems. DBMS [Otkrytyye sistemy. SUBD], 4, 14-17.

5. Aronsen, J. M., Hooft, R. W. W., Slavec, A. et al. (2021). Recommendations on FAIR metrics for EOSC. Publications Office of the European Union. URL: http://op.europa.eu/en/publication-detail/-/publication/ced147c9- 53c0-11eb-b59f-01aa75ed71a1/.

6. Blümel, I., Dietze, S., Heller, L. et al. (2014). The Quest for Research Information. Procedia Computer Science, 33, 253-260.

7. Collins, S., Genova, F., Harrower, N. et al. (2018). Turning FAIR into reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data. Publications Office of the European Union. URL: https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf.

8. Lamprecht, A.-L., Garcia, L., Kuzak, M. et al. (2020). Towards FAIR principles for research software. Data Science, 3(1), 37-59. URL: https://doi.org/10.3233/DS-190026.

9. RDA FAIR Data Maturity Model. (2020). FAIR Data Maturity Model: specification and guidelines. Research Data Alliance. DOI: https://doi.org/10.15497/RDA00050.

10. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3 (1), 160018. DOI: https://doi.org/10.1038/sdata.2016.18.

11. Michelsen, K., Brand, H., Achterberg, P. W., Wilkinson, J. R. (2015). Promoting better integration of health information systems: best practices and challenges. Copenhagen: WHO Regional Office for Europe (Health Evidence Network (HEN) synthesis report).

12. Frey, L. J. (2018). Data integration strategies for predictive analytics in precision medicine. Per Med., 15 (6), 543-551. DOI:10.2217/pme-2018-0035.

13. Saleh, S. H., Ismail, R., Ibrahim, Z., Hussin, N. (2018). Issues, Challenges and Solutions of Big Data in Information Management: An Overview. International Journal of Academic Research in Business and Social Sciences, 8 (12), 1382-1393.

14. Kanchi, S., Sandilya, S., Ramkrishna, S., Manjrekar, S., Vhadgar, A. (2015). Challenges and solutions in big data management - an overview. Presented at 3rd International Conference on Future Internet of Things and Cloud. Rome, Italy, IEEE.

15. Krishnaveni, C., Udhayakumar, U. (2018). Big data analytics: challenges and applications for social media data, presented at National Conference on Emerging Trends in Computing Technologies. Villupuram, Tamil Nadu. International Research Group.

16. Marr, B. (2017). 3 massive big data problems everyone should know about. URL: https://www.forbes.com/sites/bernardmarr/2017/06/15/3-massive-big-data-problemseveryone-should-know-about/#7df0e5606186.

ORCID:
Ozar P. Mintser: 0000-0002-7224-4886
Larysa Yu. Babintseva: 0000-0003-2753-5489